

**Studies in Protein Post-Translational Modification Using CAD and ETD
Mass Spectrometry**

Jeremy Lynn Balsbaugh
Lebanon, Pennsylvania

Bachelor of Science in Chemistry, Messiah College, 2006

A Dissertation Presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Chemistry

University of Virginia
May 2011

Donald F Hunt
M. L. H. Coon
B. J. Goldstein
David C. P. Suck
James A. H. Murray

Acknowledgments

I must first thank my advisor, Dr. Donald F. Hunt, and his longtime protégé, Dr. Jeffrey Shabanowitz, for their willingness to accept me into the Hunt laboratory and for the vast amount of help they provided in molding me into an analytical chemist. My experience in the Hunt lab has been incredible and I appreciate the mentorship that you both provided during the first stage of my scientific career.

I owe a large amount of my success to both former and current Hunt lab graduate students and post-docs. In particular, I must thank Drs. Kristie Rose, Tara Muratore-Schroeder, Namrata Udeshi, Erin Jeffery and Phil Compton for many technical lessons and advice when I first joined the lab. Their input set me on the right track for the rest of my dissertation work. I must thank Drs. Jessica Chapman, Joshua Nicklay, Sushmit Maitra and Patrick James for their extensive contributions not only during my graduate work but also for their helpful suggestions during the preparation of this dissertation. Special thanks go to Dr. Dina Bai for her monumental help in all things computer-related, of which I can claim no expertise. In addition, I need to thank Michelle English, Aaron Bailey, Dr. Andrew Norris, Jenny Cottine, Joseph Strukl, Andrew Dawdy, Lissa Anderson, Jenn Abelin and Lichao Zhang for day-to-day contributions, insightful conversations and support during my defense period. I can honestly say that I've enjoyed working with each and every one of you and I'm excited to see where life takes you after the "Hunt lab experience."

I appreciate the contributions from Drs. David Cafiso, B. Jill Venton, James N. Demas and Ian G. Macara, who collectively comprised my candidacy exam and Ph.D. defense committees.

I cannot understate the contributions by the biological collaborators for all of my projects and particularly those contained in this dissertation. I've thoroughly enjoyed learning and working with Drs. Ian Macara, David Brautigan, Gerald Hart, Emily Bernstein, Noel Lowndes and the many post-docs and graduate students from their respective labs. I need to specifically thank Dr. Emily Hall from the Brautigan lab, Dr. Ron Copeland from the Hart lab, and Dr. Christine Tooley-Schaner and Janusz Petkowski from the Macara lab. It was a pleasure working with all of you. Your contributions to my research were invaluable and I've enjoyed learning with and from you along the way.

I've garnered support from many people during my time in Charlottesville, but I specifically thank both my immediate and extended family for their prayers and support once I left Lebanon on this journey. Mom and Dad – I appreciate all you've done to instill in me the values of honesty and hard work. I'm thankful for everything you've done to help me get to this point.

Furthermore, I need to especially thank my amazing wife for her unwavering support during my graduate experience. Lauren – we've shared late night meals and "grading nights/weekends" in lab, midnight trips to Kinko's, and you've heard more than your fair share of group meetings. I appreciate your listening ear and every

other contribution you've made to my life pertaining to my graduate studies and otherwise. I'm excited to share this journey with you, wherever it may take us. I should also thank my daily writing partner, Luna. While you may not be the best conversationalist, you kept me sane during many days of solitary dissertation writing and at least pretended to listen to my practice talks.

Lastly, I need to thank the outstanding educators at Cedar Crest High School, Messiah College, and the University of Virginia for encouraging me to study chemistry and to participate in scientific research. You have instilled in me the importance of quality education and I sincerely hope to extend this to many students in the future.

Table of Contents

Table of Contents	iv
Lists of Figures and Tables	vii
Abbreviations	xiv
Abstract	xxi
 Chapter 1	
1 Introduction to the Dissertation	
1.1 Introduction	1
1.2 References	20
 Chapter 2	
2 Comprehensive analysis of phosphorylation sites in tensin 1 reveals regulation by p38 MAPK	
2.1 Abstract	22
2.2 Introduction	24
2.2.1 Cell signaling and protein phosphorylation	25
2.2.2 Focal adhesions and cell migration	28
2.2.3 Tensin 1	31
2.2.4 Protein phosphatase 1α	33
2.2.5 DLC-1	35
2.3 Materials	36
2.3.1 Reagents	36
2.3.2 Equipment & instrumentation	39
2.4 Methods	40
2.4.1 Cell culture	41
2.4.2 Affinity purification of S-tagged tensin 1 and tensin 1 F302A	41
2.4.3 Immunoblot analysis of S-tagged tensin 1	42
2.4.4 Proteolytic digestions and sample preparation for mass analysis	42
2.4.5 HPLC capillary column assembly	43
2.4.6 HPLC-MS/MS analysis	43

2.4.7	Phosphorylation analysis using differential isotopic labeling and immobilized metal affinity chromatography for phosphopeptide enrichment	45
2.4.8	Mass spectrometric data analysis	46
2.4.9	<i>In vitro</i> kinase assays	47
2.4.10	DLC-1 binding assays	48
2.4.11	Hyperosmotic sorbitol cell treatment	48
2.5	Results and Discussion	48
2.5.1	S-tagged tensin 1 purification	49
2.5.2	<i>In silico</i> proteolytic digestion	50
2.5.3	LC-MS/MS analyses of tensin 1 WT and F302A \pm calyculin A treatment	56
2.5.4	Differential isotopic labeling phosphorylation enrichment analysis	72
2.5.5	Tyrosine phosphorylation enrichment analysis of tensin 1 WT	80
2.5.6	<i>In vitro</i> kinase assays	85
2.5.7	DLC-1 binding assays	87
2.5.8	Hyperosmotic sorbitol treatment	89
2.6	Conclusions	91
2.7	References	96

Chapter 3

3	Post-translational modification analysis of homeodomain-interacting protein kinase 1	
3.1	Abstract	98
3.2	Introduction	99
3.2.1	O-GlcNAc and O-GlcNAc transferase	99
3.2.2	HIPK1	103
3.3	Materials	106
3.3.1	Reagents	106
3.3.2	Equipment & instrumentation	109
3.4	Methods	110
3.4.1	Cell culture & Flag-HIPK1 transfection	110
3.4.2	Affinity purification of Flag-tagged HIPK1	110
3.4.3	Immunoblot and autoradiography analysis of O-GlcNAcylated HIPK1	111
3.4.4	Proteolytic digestions and sample preparation for mass analysis	111
3.4.5	HPLC capillary column assembly	112
3.4.6	HPLC-MS/MS analysis	113
3.4.7	Histamine derivatization of HIPK1	115
3.5	Results and Discussion	116
3.5.1	Affinity purification of Flag-tagged HIPK1	116

3.5.2	Immunoblot and autoradiography analysis of O-GlcNAcylated HIPK1	117
3.5.3	<i>In silico</i> proteolytic digestion	118
3.5.4	CAD and ETD analysis of Asp-N HIPK1 peptides	120
3.5.5	CAD and ETD analysis of Asp-N/trypsin HIPK1 peptides	125
3.5.6	CAD analysis of Asp-N/chymotrypsin HIPK1 peptides	130
3.5.7	Histamine derivatization of HIPK1	132
3.5.8	Future work	137
3.6	Conclusions	138
3.7	References	140

Chapter 4

4	Discovery of novel protein targets of α -N-terminal methylation and N-terminal RCC1 methyltransferase	
4.1	Abstract	144
4.2	Introduction	145
4.2.1	Protein methylation by methyltransferases	145
4.2.2	α -N-methylation of protein N-termini	146
4.2.3	RCC1 α -N-methylation	147
4.2.4	N-terminal RCC1 methyltransferase (NRMT) and its substrates	149
4.3	Materials	153
4.3.1	Reagents	153
4.3.2	Equipment & instrumentation	155
4.4	Methods	156
4.4.1	Immunoblotting for me ² -PPK & me ³ -SPK in <i>Mus musculus</i> tissue	156
4.4.2	Anti-me ³ -SPK <i>Mus musculus</i> immunoprecipitation	156
4.4.3	Proteolytic digestions and sample preparation for mass analysis	157
4.4.4	HPLC capillary column assembly	158
4.4.5	HPLC-MS/MS analysis	158
4.5	Results and Discussion	160
4.5.1	Immunoblotting for α -N-methylation in <i>Mus musculus</i> tissues	160
4.5.2	Anti-me ³ -SPK <i>Mus musculus</i> heart and spleen IP	161
4.5.3	Novel LC-MS/MS methods for analysis of N-terminal methylation	162
4.5.4	Signature loss of trimethylamine under ETD fragmentation	181
4.6	Conclusions	183
4.7	References	187

List of Figures and Tables**Chapter 1**

Figure 1.1 Ion stability in a quadrupolar field.....	8
Figure 1.2 MS1 spectrum of many coeluting peptides including tensin 1 peptide, SFSAPATQAYGHEIPLR.....	11
Figure 1.3 Fragmentation by CAD results in amide bond cleavage.....	14
Figure 1.4 The proposed ETD radical-directed fragmentation mechanism results in cleavage about the N- α C bond.....	16

Chapter 2

Figure 2.1 Phosphorylated peptide binding to Fe(III) imidodiacetate IMAC POROS MC 20 resin conjugated to a solid support.....	26
Figure 2.2 Focal adhesions are protein complexes that anchor the cytoskeleton to the extracellular matrix through transmembranal, heterodimeric integrins and many scaffolding proteins.....	30
Figure 2.3 Structural domains of human tensin 1	32
Figure 2.4 Electrophoretic mobility changes under SDS-PAGE for tensin 1 WT and F302A upon phosphatase inhibitor treatment.....	35
Figure 2.5 Full length human tensin 1 WT amino acid sequence.....	49
Figure 2.6 Purification of tensin 1 F302A with calyculin A treatment from HEK293T cells.....	50
Table 2.1 <i>In silico</i> tryptic digestion of full length tensin 1 WT.....	56

Figure 2.7 Base peak chromatogram resulting from LC-MS/MS analysis of S-tagged tensin 1 F302A with calyculin A treatment.....	59
Figure 2.8 CAD MS/MS spectrum of singly phosphorylated and carbamidomethylated tensin 1 peptide, HLGSGSVVPGSPCLDR.....	63
Figure 2.9 ETD MS/MS of singly phosphorylated and carbamidomethylated HLGSGSVVPGSPCLDR.....	64
Figure 2.10 XICs of standard peptides angiotensin and vasoactive intestinal peptide used for relative quantitation.....	65
Figure 2.11 XICs and MS1 data for carbamidomethylated tensin 1 peptide HLGSGSVVPGSPCLDR.....	67
Figure 2.12 Sequence coverage map of tensin 1.....	69
Figure 2.13 Phosphorylated residue distribution on tensin 1 WT and F302A without calyculin A treatment.....	70
Table 2.2 Tensin 1 WT and F302A phosphorylated residues rescued using calyculin A treatment.....	70
Table 2.3 Relative abundances of phosphorylation in tensin 1 WT and F302A (1-740) with calyculin A treatment.....	72
Figure 2.14 D ₃ /D ₀ ratio determination for HLGSGSVVPGSPCLDR.....	75
Table 2.4 Phosphorylated tensin 1 peptides with the greatest D ₃ /D ₀ ratios.....	76
Figure 2.15 CAD and ETD MS/MS spectra of O-GlcNAcylated tensin 1 peptide, SFSAPATQAYGHEIPLR.....	78
Figure 2.16 CAD and ETD MS/MS spectra of O-GlcNAcylated tensin 1 peptide, AQFSVAGVHTVPGSPQAR.....	79

Table 2.5 Tensin 1 O-GlcNAcylation results with calyculin A treatment.....	80
Figure 2.17 Immunoblotting tensin 1 against phosphorylated pTyr using control and peroxyvanadate treated tensin 1 WT.....	81
Table 2.6 Phosphotyrosine results following treatment with peroxyvanadate and enrichment using IMAC.....	82
Figure 2.18 Distribution of tyrosine phosphorylation on peroxyvanadate-treated tensin 1 WT.....	82
Table 2.7 Complete PTM identification summary resulting from this comprehensive analysis.....	84
Figure 2.19 Phosphor-image analysis shows active phosphorylation in the presence of ATP during incubation after lysis with equal amounts of tensin 1 purified.....	86
Figure 2.20 <i>In vitro</i> effect of kinase inhibitors on tensin 1 phosphorylation..	86
Figure 2.21 Quantitation of phosphorylation reduction in tensin 1 and co-immunopurification of p38 MAPK using anti-p38 MAPK and anti-S-tag.....	87
Figure 2.22 The relationship between DLC-1 binding and tensin 1 phosphorylation.....	89
Figure 2.23 Sorbitol treatment and the effect on tensin 1 interacting proteins.....	90

Chapter 3

Figure 3.1 Protein GlcNAcylation.....	100
Figure 3.2 GlcNAcyated peptide behavior under C18 RPLC-MS/MS.....	102
Figure 3.3 Structural domains of human HIPK1.....	104

Figure 3.4 TNF α -induced HIPK1 desumoylation by SENP1 activates ASK1-JNK signaling to promote apoptosis.....	105
Figure 3.5 Amino acid sequence of Flag-tagged human HIPK1.....	116
Figure 3.6 Immunoprecipitation of Flag-HIPK1.....	117
Figure 3.7 Flag-HIPK1 is GlcNAcylated <i>in vivo</i>	118
Table 3.1 <i>In silico</i> digestion of human HIPK1 with endoproteinase Asp-N..	120
Figure 3.8 Base peak chromatogram of endoproteinase Asp-N-generated HIPK1 peptides.....	122
Figure 3.9 Proteolysis of HIPK1 using endoproteinase Asp-N results in only 50% sequence coverage using CAD and ETD MS.....	123
Figure 3.10 HIPK1 peptide DQGLLLPAPAVEHIVVTAA was found to be GlcNAcylated based on accurate mass and CAD MS/MS.....	124
Figure 3.11 ETD MS/MS spectrum of singly GlcNAcylated peptide DQGLLLPAPAVEHIVVTAA.....	125
Table 3.2 Summary of HIPK1 PTMs identified from CAD and ETD analysis of the initial Asp-N digestion.....	125
Figure 3.12 Proteolysis using Asp-N and trypsin sub-digestion results in nearly 80% HIPK1 sequence coverage.....	126
Figure 3.13 ETD MS/MS spectrum of singly GlcNAcylated, carbamidomethylated HIPK1 peptide, DcTVATQASGLLSNK.....	127
Table 3.3 Summary of PTMs identified following HPLC-MS/MS analysis of Asp-N/trypsin digested HIPK1.....	128

Figure 3.14 Charge state distribution of singly GlcNAcylated HIPK1 peptide VHQPVSVGPSLL.....	131
Figure 3.15 CAD MS/MS of GlcNAcylated HIPK1 peptide VHQPVSVGPSLL results in typical signature GlcNAc ion losses and minimal fragment ions..	132
Figure 3.16 EDC-mediated peptide derivatization with histamine.....	134
Figure 3.17 ETD MS/MS spectrum of singly GlcNAcylated and histamine-derivatized HIPK1 peptide VHQPVSVGPSLL.....	135
Table 3.4 Summary of PTMs identified on HIPK1 using multiple proteases and histamine-derivatization.....	136

Chapter 4

Figure 4.1 N-terminal methylation of GFP-tagged RCC1 regulates interactions with chromosomes and mediates spindle pole organization.....	148
Figure 4.2 Full methylation of protein N-termini leads to a fixed positive charge on the α -amino group.....	149
Figure 4.3 Interrogation of NRMT structure and specificity.....	151
Table 4.1 Probable NRMT substrates based on SPK, APK, PPK and PK N-terminal sequences following initiating Met cleavage.....	153
Figure 4.4 Immunoblotting against me ² -PPK and me ³ -SPK in various <i>Mus musculus</i> tissue.....	161

Figure 4.5 SDS-PAGE and Coomassie staining identifies proteins from spleen and heart tissue me ³ -SPK IPs.....	162
Table 4.2 N-terminal peptides generated from proteolysis of probable NRMT substrates.....	164
Figure 4.6 Novel strategy for enrichment of N-terminal peptides using methylation-specific antibodies and MWCO filters.....	166
Figure 4.7 ETD MS/MS spectrum of N-terminally trimethylated MLC, regulatory B-like N-terminal peptide SSKRAKTKT ³ KKRPQRATSNVGAMGDQSQIQEFKE.....	168
Figure 4.8 Three unique MLCs in varied abundance were observed to contain N-terminal SSK trimethylation.....	169
Table 4.3 Summary of peptide and source proteins identified with N-terminal di- or trimethylation.....	172
Figure 4.9 ETD MS/MS spectrum of the N-terminal peptide of mouse RCC1 modified with N-terminal dimethylation.....	173
Figure 4.10 ETD MS/MS spectrum of the N-terminal peptide of SET modified with N-terminal trimethylation.....	175
Figure 4.11 ETD MS/MS spectrum of the N-terminally trimethylated peptide from mouse ribosomal protein L23a.....	176

Figure 4.12 ETD MS/MS spectrum of the N-terminal peptide of kelch-like protein 31 modified with N-terminal trimethylation.....	178
Figure 4.13 ETD MS/MS spectrum of the N-terminal peptide of myosin light polypeptide 3 modified with N-terminal trimethylation.....	179
Figure 4.14 ETD MS/MS spectrum of the N-terminal peptide of myosin light polypeptide 2 modified with N-terminal trimethylation.....	180
Figure 4.15 SDS-PAGE and western blot analysis of anti-me3-SPK immunopurified proteins.....	181
Figure 4.16 Secondary electron transfer to multiply charged c-ions directs fragmentation at the N-terminal trimethylamine.....	182

Abbreviations

ABD	actin-binding domain
AGC	automated gain control
AIP1	ASK1-interacting protein
Ala, A	alanine
amu	atomic mass unit
Arg, R	arginine
ASK1	apoptosis signal-regulating kinase 1
Asn, N	asparagine
Asp, D	aspartic acid
ATP	adenosine triphosphate
AV	number of spectra averaged
CAD	collision-activated dissociation
CI	chemical ionization
CID	collision-induced dissociation
Cys, C	cysteine
Da	Dalton, 1 amu
DC	direct current
DLC-1	deleted in liver cancer-1
DMSO	dimethylsulfoxide

DNA	deoxyribonucleic acid
ECD	electron capture dissociation
ECM	extracellular matrix
EDC	(N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride)
EDTA	ethylenediaminetetraacetic acid
EM	electron multiplier
ESI	electrospray ionization
ETD	electron transfer dissociation
FAB	focal adhesion binding domain
FAK	focal adhesion kinase
FETD	front-end electron transfer dissociation
fmol	femtomole (10^{-15} mol)
FT	Fourier transform
Gal	galactose
GlcNAc	<i>N</i> -acetyl-D-glucosamine
Gln, Q	glutamine
Glu, E	glutamic acid
Gly, G	glycine
gSer, gS	GlcNAcylated serine
gThr, gT	GlcNAcylated threonine

GTPase	guanosine triphosphate hydrolase
HIPK1	homeodomain-interacting protein kinase 1
His, H	histidine
HPLC	high performance liquid chromatography
ICR	ion cyclotron resonance
Ile, I	isoleucine
IMAC	immobilized metal affinity chromatography
in	inch
IP	immunoprecipitation
ITMS ⁿ	ion trap tandem mass spectrometry
i.d.	inner diameter
I2PP2A	phosphatase 2A inhibitor
kDa	kiloDalton (10^3 Da)
Leu, L	leucine
LTQ	linear trap quadrupole
Lys, K	lysine
L	liter
m	mass
m	meter
M	molarity (mole/L)

MAPK	mitogen-activated protein kinase
MDa	megadalton (10^6 Da)
me	monomethyl
me ₂	dimethyl
me ₃	trimethyl
Met, M	methionine
MHC	myosin heavy chain
MLC	myosin light chain
Mox	oxidized methionine
mRNA	messenger RNA
MS	mass spectrometry
MS ⁿ	tandem mass spectrometry
MW	molecular weight
m/z	mass-to-charge
nM	nanomolar (10^{-9} M)
NMR	nuclear magnetic resonance
NL	normalized ion counts
NRMT	N-terminal RCC1 methyltransferase
OMSSA	open mass spectrometry search algorithm
O-GlcNAc	O-linked N-acetylglucosamine

OGT	<i>O</i> -GlcNAc transferase
NRMT	N-terminal RCC1 methyltransferase
o.d.	outer diameter
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
Phe, F	phenylalanine
pmol	picomole (10^{-12} mol)
PP1	protein phosphatase 1
Pro, P	proline
pSer, pS	phosphorylated serine
PTB	phosphotyrosine binding domain
pThr, pT	phosphorylated threonine
PTM	post-translational modification
QIT	quadrupolar ion trap
QLT	quadrupolar linear ion trap
PPM	parts per million
RCC1	regulator of chromatin condensation 1
RF	radiofrequency
RhoGAP	Rho GTPase-activating protein

RP	reverse phase
RNA	ribonucleic acid
RT	retention time
SDS	sodium dodecyl sulfate
Ser, S	serine
SET	patient SE translocation
SUMO	small ubiquitin-like modifier
TAF-1	template-activating factor 1
TBS	tris-buffered saline
Thr, T	threonine
Thx	thioredoxin
TLC	thin-layer chromatography
TNF	tumor necrosis factor
Trp, W	tryptophan
Tyr, Y	tyrosine
TLC	thin layer chromatography
tRNA	transfer RNA
UDP	uridine diphosphate
µg	microgram (10 ⁻⁶ g)
µL	microliter (10 ⁻⁶ L)

Val, V valine

WT wild-type

Abstract

Proteins are macromolecules that perform specialized and critical functions for nearly every biological process. Regulation of protein function is required to prevent diseases including cancer. Covalent post-translational modifications (PTMs) can alter protein structure to regulate the activity of many proteins. While several methods are currently used to identify PTMs, mass spectrometry (MS) stands at the forefront because it is an unbiased and sensitive technique that can provide unambiguous identification of the 200+ PTMs currently known. Described within this dissertation are three applications of high resolution MS that implement high performance liquid chromatography, collision-activated dissociation (CAD) and electron transfer dissociation (ETD) to identify novel PTMs on various proteins.

First, we describe a comprehensive phosphorylation site analysis of human tensin 1, an extremely large scaffolding protein required for controlled cell migration. By implementing optimized proteolysis, CAD and ETD MS, and phosphopeptide enrichment, we identified an incredible 72 phosphorylation sites on tensin 1. Using differential isotopic peptide labeling with phosphopeptide enrichment, we identified potential protein phosphatase 1 α dephosphorylation target sites. This study also revealed regulation by p38 α MAPK, a Ser/Thr kinase that extensively phosphorylates tensin 1.

Second, we demonstrate the utility of ETD for the analysis of *O*-linked N-acetylglucosamine (*O*-GlcNAc), a monosaccharide PTM that is notoriously difficult to study using MS. By implementing high resolution MS, ETD and recently developed

peptide charge state modification chemistry, we identified 3 novel *O*-GlcNAc and 15 novel phosphorylation sites on human homeodomain-interacting protein kinase 1.

Third, we present new MS-based methods for the characterization of α -N-methylation. Although methylation of Lys and Arg side chains has been extensively studied, very few mammalian substrates for α -N-methylation are known. Our new enrichment method incorporates α -N-methylation-specific antibodies, molecular size filtering, high resolution MS and ETD fragmentation. Application of these methods to mouse heart and spleen tissue lysates afforded the identification of several N-terminally methylated proteins that were previously unknown. In addition, we expanded the known N-terminal consensus sequence for mammalian α -N-methylation to include Ala/Pro/Ser-Pro-Lys and Ser-Ser-Lys. As a result, we provided a significant increase in the number of known mammalian α -N-methylation targets.

1 Introduction to the Dissertation**1.1 Introduction**

The central dogma of molecular biology first introduced by Francis Crick in 1958 describes the general flow of organismal hereditary information through three separate macromolecules (1, 2). This model dictates that the genome, or the collective gene population encoded by DNA, can be transcribed into another nucleic acid-based structure via a process termed transcription to yield the transcriptome, or the RNA content of a cell. The transcriptome consists of three types of mature RNA all of which are required for protein synthesis: ribosomal (rRNA), messenger (mRNA), and transfer (tRNA). Proteins, which are synthesized as dictated by the sequence information originating in DNA that is later incorporated into RNA, are composed of amino acid chains that form large macromolecules up to 1.7 MDa. The proteome, or the entire protein population within a cell, carries out extensive and essential cellular processes including transcription and translation through individual and concerted efforts with other macromolecules (3).

Proteomics is a field of study that focuses on the characterization of individual protein expression, structure and function (4). Regulation of critical biological processes is carried out in large part by proteins, which drives the desire to understand their roles particularly for efforts leading to diagnosis and disease treatment options. The complex proteome arises from the unique amino acid primary sequences that make up proteins. This diversity is made possible by the

unique conjugation of 20 unique amino acids available for synthesis in mammalian cells. Higher order protein structure results from this primary sequence. Because specialized protein function relies on specific structure, the primary sequence is also directly related to function. The entire 25,000 gene-containing human genome has been sequenced and allowed for the elucidation of these sequences. This information has guided current research efforts away from methods designed solely for protein sequencing and toward deciphering protein function and corresponding regulatory mechanisms.

The functional repertoire of the proteome is further expanded by over 200 known different post-translational modifications (PTMs) (5, 6). These essential modifications not only promote diversity but also allow for regulation of signaling pathways during inter- and intra-cellular communication. Most PTM events are reversible and involve addition and removal of small chemical moieties (ie. phosphate, monosaccharide and methyl groups) covalently attached to side chains of specific amino acid residues within a folded protein, yet some modifications have also been found to involve extreme protein termini as well (6-8). PTM can also result in the addition of large moieties including entire proteins, polysaccharide chains, lipids or the removal of amino acids due to proteolytic cleavage (9, 10). Protein functional roles include recognition of PTMs and/or catalysis leading to the physical attachment or detachment of PTMs on residues. This specificity allows for tight regulation of signaling pathways necessary for cell-to-cell or organelle-to-organelle communication (9). This is achieved using the ability to transmit signals

quickly through the cell in order to activate or deactivate processes required for survival. Reversible phosphorylation, one of the first PTMs discovered, allows intracellular signaling by the enzymatic addition of phosphate moieties onto serine, threonine, and tyrosine side chains in mammalian cells. Protein phosphorylation, which is estimated to occur on approximately one-third of all proteins, continues to be a major research focus due to known regulatory roles in cell cycle progression, DNA damage response and cellular growth (6, 11).

Attempts at PTM characterization in the past have mainly implemented indirect methodologies such as Edman degradation and TLC-based methods (6). These methods are disadvantageous due in part to the requirement for large amounts of protein and the inability to accurately identify novel modifications. Alternate methods have involved the use of radiography techniques in order to identify types of PTM incorporated into proteins (12). These methods have hazards and struggle with identification and unambiguous localization of specific PTMs on individual amino acid residues. Likewise, antibody-based methods have been used to localize, identify and enrich post-translationally modified proteins *in vivo* yet are unable to confidently localize PTMs to specific residues due to cross reactivity and nonspecific binding tendencies.

The observation that essentially all covalent PTMs alter the molecular structure, and therefore the molecular weight (MW), of the modified protein has prompted mass spectrometry (MS) to rise to the forefront of PTM characterization

in the proteomics field. MS allows the ability to accurately measure masses of proteins and proteolytically-generated peptides, including those that contain PTMs. Recent developments in mass spectrometric technologies allow unambiguous site identification of PTMs, which are imperative for biochemical functional studies and thorough understanding of signaling pathways.

The use of analytical MS for proteomics has progressed extensively in the last several decades. Mass spectrometers exploit the ability to manipulate gas phase ions through time and space using both DC and RF fields under vacuum. As a result, mass spectrometers are able to measure masses, or more precisely mass-to-charge (m/z) ratios, for ionized molecular species. For many years, the physical introduction of large peptide and protein cations into mass spectrometers kept the field of proteomics at bay. However, with the groundbreaking invention of electrospray ionization (ESI) by Fenn et al. in 1985, peptide cations and other large biomolecules including whole proteins can now be ionized efficiently at atmospheric pressure directly into mass spectrometer inlets (13). This sensitive, “soft” ionization technique, which keeps multiply protonated molecular species intact upon ionization, allows the direct introduction of solubilized peptides and proteins into the instrument for mass analysis. Using capillary flow, molecular species can be ionized through the application of a DC potential directly to the capillary and solvent system. Multiply charged droplets containing the molecular species of interest are emitted from the tip of the capillary column into the instrument ion source. Upon desolvation assisted by a heated bath gas or heated

stainless steel capillary, multiply charged droplets condense to result in multiply charged gas phase ions, $(M+nH^+)^{n+}$ (14). It is common for larger ionized peptides and proteins to exist simultaneously in a series of consecutive charge states, as dictated by amino acid composition, degree of ion basicity and ESI solvent composition. The advent of ESI has allowed for the ionization and analysis of biomolecules previously unable to be characterized due to the ability to bring species of variable mass (100 – 100,000 Da) to m/z values easily measured by current ion trap instrumentation ($<2,000 m/z$).

Additionally, ESI has permitted the direct integration of high performance liquid chromatography (HPLC), a powerful tool for the separation of complex mixtures of proteins and peptides, with mass spectrometry. Due to the great complexity of most biological samples, HPLC is required for fractionation of proteins and peptides prior to mass analysis. The online coupling of HPLC and mass spectrometry form the backbone of the majority of recent proteomic studies.

Fractionation of proteolytic peptides typically follows digestion, wherein solubilized species are separated based on interactions with a stationary phase. Reverse-phase (RP) HPLC conditions are commonly used and require a polar mobile phase and nonpolar stationary phase, usually consisting of C18-based silica material. The created molecular environment is suitable for separation of molecular species based on increasing hydrophobicity. This powerful separation technique promotes higher levels of dynamic range due to the ability to separate, and

therefore simplify, biologically complex samples for mass analysis over an extended period of time for the identification of thousands of peptides within a single experiment. Alternatively, a technique dubbed “top down” proteomics strives to characterize intact proteins. Limitations in protein separation efficiencies characteristic of current chromatographical methods hold back the success of these analyses. Ultimately, it is the implementation of ESI with ion trap technology that drives success in both protein and peptide-based proteomic studies.

Ion trap mass analyzers are well suited for MS experiments due to their high sensitivity, unit resolution, wide mass range (0 – 2,000 m/z and greater) and fast scan speeds (14). Scan speeds on the order of milliseconds allow for rapid data acquisition of species separated in real time using HPLC methods. These performance characteristics are based on the capability of ion trap mass spectrometers to transport, isolate and selectively analyze peptide cations quickly using ion stability characteristics rooted in basic quadrupolar theory. At its most basic level, a quadrupole is a set of four parallel metal rods placed 90 degrees apart (**Figure 1.1A**). To each set of opposing rods a prescribed potential is applied, ϕ_0 , represented by a combination of DC (U) and RF fields ($V \cos(\omega t)$, where V represents RF amplitude, ω the angular frequency related to frequency in Hz and t , time). The collective electric field is represented by **Equation 1.1** (14-16).

$$(1.1) \quad \phi_0 = U - V \cos(\omega t), \quad \text{where } \omega = 2\pi f$$

With the application of equal potentials to opposite rods and potentials 180 degrees out of phase for neighboring rods, ions are trapped or selectively transported through the quadrupole. As can be expected, the spacing between quadrupolar rods directly affects the oscillating electric field applied to analyte ions. The two dimensional applied potential, $\phi_{(x,y)}$, experienced by analytes at any point within the quadrupole can be determined by **Equation 1.2**, where ϕ_0 represents the applied potential, r_0 the radius between rods, x and y the distances from the center of the quadrupolar field (14-16).

$$(1.2) \quad \phi_{(x,y)} = \phi_0 \frac{x^2 - y^2}{r_0^2}$$

In this fashion, the motion of an analyte ion through a quadrupolar field can be described through a quadratic equation commonly known as the Mathieu equation, **Equation 1.3**, whereas u represents the displacement in x and y directions, ξ is equal to $\omega t/2$ with two dimensionless variables a and q (14-16).

$$(1.3) \quad \frac{d^2 u}{d^2 \xi} + (a_u - 2q_u \cos(2\xi))u = 0$$

Variables a and q describe the stability of ions within a quadrupolar field as represented by the first stability diagram (**Figure 1.1B**).

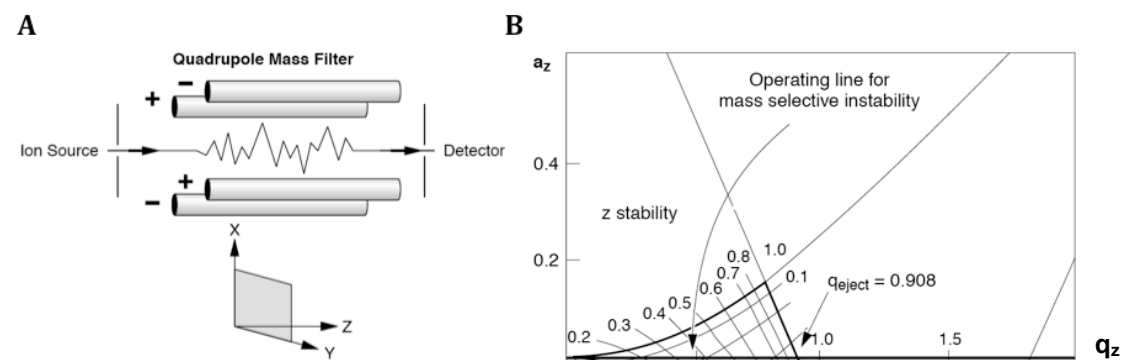


Figure 1.1 Ion stability in a quadrupolar field. A) Opposite quadrupoles carry identical potentials while neighboring rods carry opposite potentials. B) Stability in the z -direction in a quadrupolar field is represented by ions contained within the shaded region and is determined by the dimensionless parameters a and q in the Mathieu equation. Since DC potentials are removed in most quadrupolar fields, ion stability is governed by q . All ions are ejected from the trap at $q = 0.908$. Because ions of different mass and charge are stable at unique ranges of q , ions can be isolated and/or ejected either selectively or simultaneously. Images reproduced from (17).

The numeric values of a and q are determined by **Equation 1.4**, whereas e represents particle charge, and m particle mass; these parameters dictate which ions of varying m/e , or m/z , are stable within the quadrupolar field (14-16). Any ions outside of the stability region will not be trapped and will either exit the quadrupole through apertures between rods toward an ion detector or simply collide with the rods and escape detection.

$$(1.4) \quad a = \frac{8eU}{m\omega^2 r_0^2} \quad , \quad q = \frac{4eU}{m\omega^2 r_0^2}$$

It should be noted that because parameter a is directly proportional to the applied DC potential (U), which in most quadrupolar instruments is not applied, it is equal to zero. This suggests that ion stability is governed by q , which is directly proportional to the applied RF potential (U) and inversely proportional to m/z . Therefore, ions of varying m/z values can be selectively isolated or transported for mass analysis by

varying q to values represented on the x-axis of **Figure 1.1B** (15). Values of $q > 0.908$ result in complete ion instability. It is the ability to isolate and destabilize ions directly into detectors (ie. electron multipliers) using mass selective instability that allows for sequential mass analysis of both precursor and product ions.

The development of an ion trap applied the basics of quadrupolar theory to a three-dimensional quadrupolar ion trap (QIT) for mass selective ion storage followed by ion detection (16). In contrast to the quadrupolar rod array previously described, a QIT implements two end-cap electrodes surrounding a single ring electrode (18). This instrumentation is still in use today and allows trapping of narrow m/z windows to isolate ions for mass analysis. Ions are detected by ejection through apertures in the trap into electron multipliers. A main disadvantage of this instrument is the limitation of the number of ions that can be trapped simultaneously within the QIT. To circumvent this issue, the quadrupolar linear ion trap (QLT) was designed to physically extend the trap in the z -direction to increase the number of ions trapped during mass analysis. QLT instruments are widely used in proteomic analyses today due to fast scan speeds (10 scans/s) in addition to femto-, atto- and zeptomolar detection sensitivities (12). Such high sensitivities are achieved through the use of electron multiplier (EM) detectors. EM detectors function on the principal of signal amplification initiated by an ion beam selectively ejected from the ion trap colliding with a metal oxide-coated (MbO) dynode typically held at a high potential ($>\pm 1000$ kV). Emission of secondary electrons resulting from the initial ion beam-dynode collision are directed into similar, subsequent

collisions and ultimately allow signal gain values greater than 10^5 (14). The resulting signal intensities are directly proportional to the number of ions initially ejected from the ion trap. Signal gain of this magnitude allows robust detection of very few ions and makes EMs invaluable for low abundance analyte detection.

While QLT instruments have many benefits including high sensitivity and fast scan cycles, they do suffer from limited resolution. This can be circumvented through integration with high resolution ($\sim 100,000$) mass analyzers, leading to QLT-hybrid mass spectrometers. The Fourier transform-ion cyclotron resonance (FT-ICR) mass analyzer consists of a small cubical or cylindrical cell placed within a static magnetic field. Ions are trapped within the cell by trapping plates and oscillate at a frequency proportional to its m/z ratio as a result of the experienced magnetic field (14, 19). Similarly, ions within an Orbitrap mass analyzer oscillate at a m/z -dependent frequency in a purely electrostatic field (20, 21). Both analyzers detect the image current generated by the unique oscillation of ions as dictated by their m/z values. The complicated time-domain signal is converted to a frequency-domain signal using Fourier transform. Because the frequency of oscillating ions can be measured very precisely, high mass accuracy measurements within 1 ppm or less result (14, 19). This can result in routine resolution in excess of 100,000 and monoisotopic mass resolution. Due to small differences in certain amino acid and PTM masses (< 1 Da), highly accurate mass measurements are imperative for confident precursor ion identifications. In addition, computer software designed to aid peptide identification benefits greatly from accurate mass data by decreasing

data search times when using small mass tolerances. Due to the benefits of high mass accuracy, most proteomics experiments use high resolution mass analyzer-QLT hybrid instruments like the LTQ-Orbitrap and LTQ-FT-ICR (Thermo Scientific).

Mass analyses using high resolution hybrid ion trap mass spectrometers generally begin with an initial survey scan (full MS or MS¹) over a prescribed m/z range in which all HPLC-fractionated and ESI-generated precursor ions are analyzed simultaneously in either the Orbitrap or ICR cell. Each ion's respective m/z values are determined and reported in a single mass spectrum integrating measured ion intensities (**Figure 1.2**). Although accurate precursor masses can be helpful for the identification of intact peptides, unambiguous identifications can only be afforded using sequence information gained from the fragmentation of precursor ions during tandem MS events.

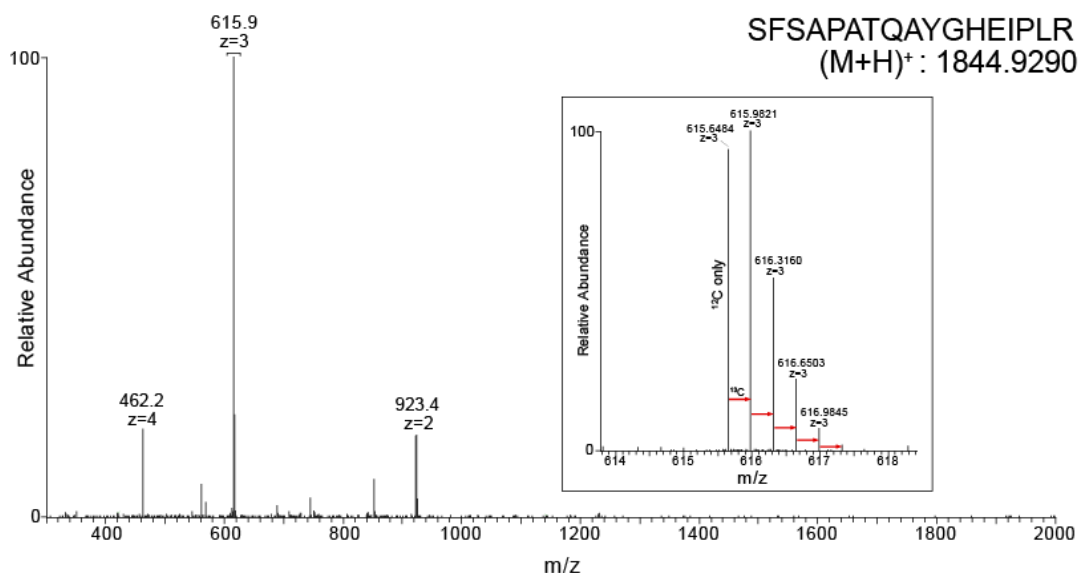


Figure 1.2 MS¹ or full MS spectrum of many coeluting peptides including tensin 1 peptide, SFSAPATQAYGHEIPLR. All ions are reported in the MS¹ spectrum with intensities proportional to the number of ions trapped and detected. The most abundant peptide in this spectrum is SFSAPATQAYGHEIPLR (tensin 1).

Positive ESI promotes efficient ionization of basic peptides and commonly results in ions within a range of charge states as seen above (+2 to +4). Inset: Expansion of 614-818 m/z region displays the ability to achieve monoisotopic resolution using high resolution analyzers including the Orbitrap and FT-ICR cell.

Tandem mass spectrometry (MS^n) refers to the practice of sequential MS events, which occur post-ionization and involve the activation of analyte ions in order to induce the generation of a series of informative fragment ions (22). Fragmentation techniques used in proteomics aim to dissociate analytes along the peptide backbone to generate fragment ions that cover each amino acid residue of the precursor peptide or protein. Backbone fragmentation between each amino acid in any peptide ideally yields complementary fragment ions; that is, a series of ions that include each of the peptide termini. Tandem MS techniques currently employed in proteomics generally rely on one of two methods of ion activation: either peptide collisions with a neutral gas or electron capture by peptide or protein cations.

For many years, collision-activated dissociation, also referred to as collision-induced dissociation (CAD and CID respectively), was the fragmentation method of choice for all MS-based proteomics. However, while still widely used today, CAD is now considered by many to be a complementary technique to those more recently developed. Under low energy (<200 eV) CAD conditions used in virtually all ion trapping mass spectrometers, peptide cations selected for MS/MS analysis are first isolated using appropriate m/z-specific RF potentials and are then subjected to collisions with an inert gas (He, Ar) (14, 22). Due to the nature of each low energy collision, single collisions do not normally result in peptide fragmentation. Instead,

sequential collisions with large, inert atoms over tens of milliseconds encouraged by supplemental RF-induced ion activation result in the transfer of additional translational energy to the peptide cation (22, 23). The translational energy imparted to peptide cations via collisions is converted to internal vibrational energy (14, 22). Once the vibrational energy within a peptide cation surpasses the energy threshold for bond dissociation, it will preferentially cleave at the amide bond to result in the generation of two fragment ions (**Figure 1.3**) (14, 22). These fragment ions include one of the original peptide N- or C-termini and are named accordingly; b- ions include the intact peptide's N-terminus whereas y-ions include the original C-terminus. Fragment ions are not excited by the precursor m/z-specific supplemental RF voltages and are not energetically encouraged to undergo collisions or extensive secondary fragmentation.

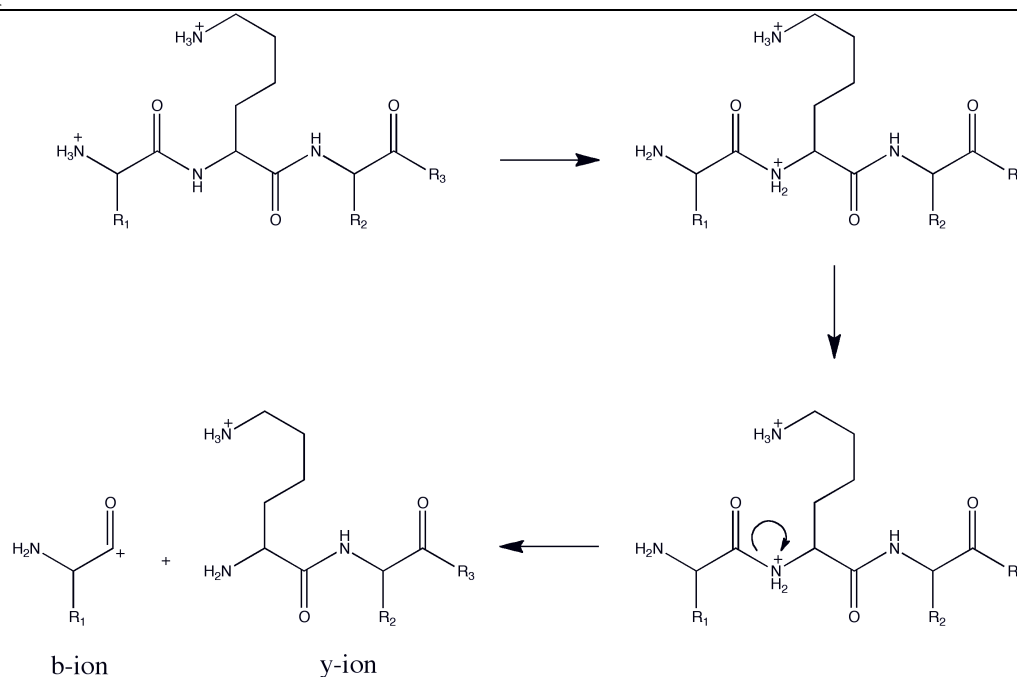


Figure 1.3 Fragmentation by CAD results in amide bond cleavage. This produces complementary b- and y-type ions which include the N- and C-terminus, respectively.

Optimally, the resulting MS/MS spectrum consists of a series of equally abundant fragment ions corresponding to cleavage at all amide bonds contained within the analyte. Peptide sequences can be deduced because the mass difference between consecutive b- and y-ions corresponds to the residue mass of a single amino acid. Therefore, by calculating mass differences between consecutive b- and y-ions, amino acid sequences can be determined. Peptide sequencing using MS/MS spectra in this manner is referred to as *de novo* sequencing and is commonly used for unknown sequence elucidation. Assignment of fragment ions in any MS/MS spectrum, which is made easier by complete and random fragmentation along the entire backbone, allows for confident sequence assignment of the dissociated precursor ion.

CAD is optimal for efficient dissociation of short (< 15 residues), low charge ($z=+2$), tryptic peptides due to relatively equal preferences for backbone cleavage throughout the molecule. However, longer (>15 residues) and more highly charged ($z\geq+3$) peptide cations dissociate poorly under CAD conditions due to the lack of random backbone protonation, which is the driving force behind the CAD mechanism (24, 25). Another disadvantage of CAD extends to peptides modified with labile PTMs such as phosphate and *O*-GlcNAc. The high degree of lability for these chemical moieties under gas phase CAD conditions leads to preferential cleavage at the PTM-to-side chain linkage, resulting in a severe lack of peptide sequence information. This is evident in MS/MS spectra displaying a single, prominent ion corresponding to the intact peptide less the PTM mass. Because the end goal of MS-based proteomic studies is the ability to sequence peptides to identify proteins and to localize PTMs to specific residues, this property of CAD is unfavorable. Therefore, efforts have focused on the development of additional fragmentation techniques for peptide dissociation.

In 1998, McLafferty and coworkers set in place the groundwork for the second type of peptide fragmentation in use today rooted in mechanisms of radical-directed dissociation. Electron capture dissociation (ECD) promotes fragmentation resulting from the capture of near-thermal electrons in the Penning trap of an FT-ICR mass spectrometer (26). Capture of an electron by a peptide cation directs fragmentation through a process that is independent of molecular vibrational excitation. Therefore, this exothermic and nonergodic process produces random

fragmentation along the backbone of peptides with minimal influence by

neighboring amino acid content, which has been shown to be a disadvantage using CAD (27). Importantly, because ECD is independent of vibrational excitation, PTMs are retained upon peptide dissociation (24). This allows for unambiguous site identification of moieties such as phosphate. Fragmentation resulting from electron capture causes bond cleavage at the N- α C bond to create complementary c- and z \bullet -ions, which include the N- and C-terminal residues, respectively (**Figure 1.4**). ECD-generated MS/MS spectra can be used to sequence peptides and proteins similarly to CAD sequencing once the mass differences between the two types of fragment ions are considered.

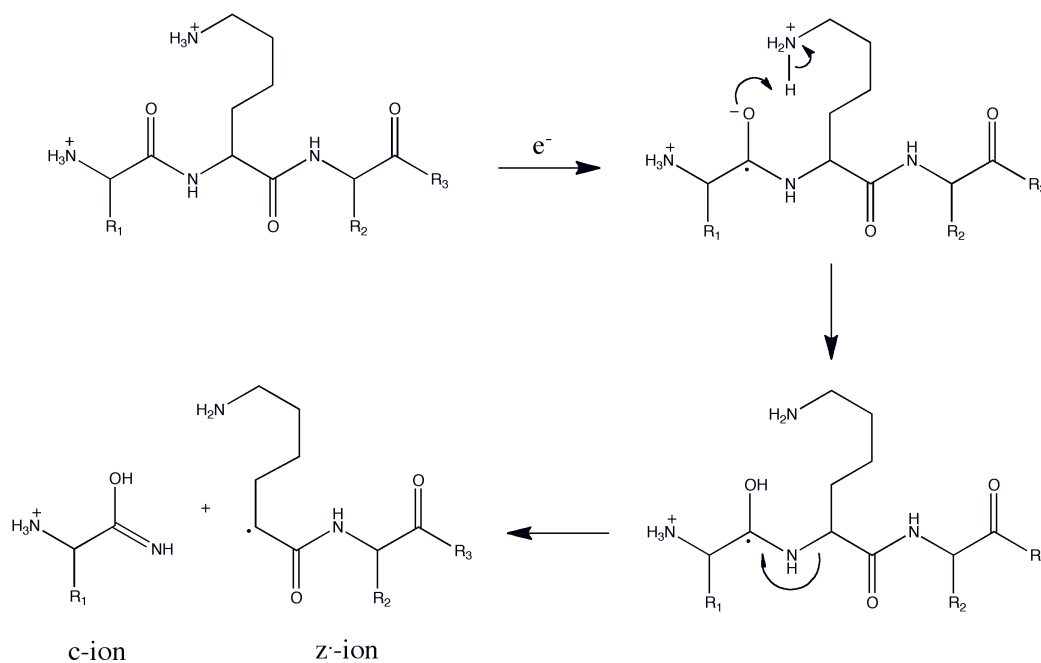


Figure 1.4 The proposed ETD radical-directed fragmentation mechanism results in cleavage about the N- α C bond. Dissociation about this bond results in complementary c- and z \bullet -type ions which include N- and C-terminus, respectively. Figure adapted from (22).

While ECD has opened the door for proteomics to successfully dissociate and sequence highly charged peptide and intact protein cations, it has been limited to mainly FT-ICR instruments, which can be financially impractical to purchase and maintain. ECD on FT-ICR instruments involves immersion of peptide cations in an excess of near-thermal electrons within the Penning trap (25). Recently, a time-of-flight instrument from Hitachi has also been modified to perform ECD (28). In contrast to these instruments, ion trap mass spectrometers are much more widespread in the proteomics community and isolate peptide cations using RF fields. Thermal electrons cannot be effectively trapped within RF fields well enough to perform ECD efficiently (21). In addition, ECD spectra of high quality are normally represented as averaged scans taken over several minutes (29). This precludes analysis on a chromatographic timescale, which is essential for most high throughput proteomic analyses and characteristic of the workhorse nature of QLT-based MS experiments.

In 2004, Syka et al. developed an analogous ion-ion technique termed electron transfer dissociation (ETD) to circumvent these issues. ETD is QLT-friendly and results in similar fragmentation patterns to ECD with a major mechanistic difference. Radical fluoranthene anions generated in an external chemical ionization (CI) source serve as electron donors to multiply charged peptide cations to replace the need for an abundance of thermal electrons (25). Modifications to existing Thermo LTQ technology allows for ETD reactions on millisecond timescales, which is appropriate for online chromatographic separations. Briefly, the three

segmented linear ion trap present in Thermo LTQ instruments allows for the selective and simultaneous isolation of peptide cations and radical anions using appropriate DC and RF potentials, after which both are allowed to react within the trap once DC potentials are dropped (25). Once the ETD reaction proceeds, all fragment ions are retained in the trap for mass analysis.

ETD, which ultimately leads to fragmentation patterns essentially identical to that of ECD, results in c- and z•-ion-containing spectra characteristic of random fragmentation along the peptide backbone and preserves labile PTMs for efficient site localization (25, 29). Performance evaluations have determined that ETD is optimal when applied to multiply charged peptide cations where $z \geq +3$ and overall charge density is high (300 to 800 m/z) (24). Currently, the application of ETD has extended from stand alone QLTs to LTQ-Orbitrap hybrid instruments with back end CI sources for ETD reagent generation. This instrument allows for acquisition of high resolution MS and MS/MS spectra using CAD and ETD for peptide interrogation (30). Additionally, front end ETD (FETD) is a recent technique developed in the Hunt laboratory that allows ETD to be implemented on any instrument including high resolution LTQ-Orbitrap or LTQ-FT-ICR MS instruments following in-house modifications to the front end ESI source (31). This technology eliminates the need for a large CI source for ETD reagent generation by adding reagent gas flow and glow discharge capabilities to a modified ESI front-end ion source. This source can generate radical anions of fluoranthene or azulene to facilitate electron transfer.

Because the in-house instrumentation modification affects only the front-end ESI

source, ETD can theoretically be a modification for any current QLT instrument.

Research described herein reflects ETD data acquired using FETD instrumentation.

Although ETD is recommended for studies using MS for the identification of labile PTMs, it is by no means reserved as an exclusive tool. The preference of CAD for peptides of low charge can be exploited for the mass analysis of some similarly charged post-translationally modified peptides. The most comprehensive MS studies compile both CAD and ETD data for protein and PTM characterization. These complementary techniques are especially important for analyzing proteolytic digestions resulting in peptides of variable charge and length. This variability results due to unique amino acid sequences in proteins that are not well suited for digestion by a single protease. It is for this reason that research described within this dissertation uses both CAD and ETD MS for the elucidation of PTMs on proteins.

This dissertation outlines the implementation of HPLC peptide separations, CAD and ETD mass spectrometry and recently developed peptide charge modification strategies for the characterization of many novel phosphorylation and *O*-GlcNAcylation PTMs on human tensin 1 (Chapter 2) and homeodomain-interacting protein kinase 1 (Chapter 3). In addition, we present the identification of novel substrates for α -N-terminal methylation catalyzed by the newly discovered N-terminal RCC1 methyltransferase (Chapter 4).

1.2 References

1. Central Dogma of Molecular Biology, Crick, F., Nature **1970** (227): 561-563.
 2. Revisiting the Central Dogma in the 21st Century, Shapiro, J. A., Annals of the New York Academy of Sciences **2009** (1178): 6-28.
 3. Posttranslational Modification of Proteins: Expanding Nature's Inventory, Walsh, C. T. **2006**, Roberts and Company Publishers, Englewood, CO 490.
 4. Proteome and proteomics: New technologies, new concepts, and new words, Anderson, N. L. & Anderson, N. G., Electrophoresis **1998** (19): 1853-1861.
 5. Proteomic analysis of post-translational modifications, Mann, M. & Jensen, O. N., **2003** (21): 255-261.
 6. Farley, A. R. & Link, A. J. (2009) in *Methods in Enzymology*, ed. Deutscher, R. R. B. a. M. P. (Academic Press): 725-763.
 7. N-terminal [alpha]-methylation of RCC1 is necessary for stable chromatin association and normal mitosis, Chen, T., Muratore, T. L., Schaner-Tooley, C. E., Shabanowitz, J., Hunt, D. F. & Macara, I. G., Nature Cell Biology **2007** (9): 596-603.
 8. NRMT is an N-methyltransferase that methylates RCC1 and retinoblastoma protein, Schaner Tooley, C. E., Petkowski, J. J., Muratore-Schroeder, T. L., Balsbaugh, J. L., Shabanowitz, J., Sabat, M., Minor, W., Hunt, D. F. & Macara, I. G., Nature **2010** (466): 1125-1128.
 9. Post-translational modifications in signal integration, Deribe, Y. L., Pawson, T. & Dikic, I., Nature Structural & Molecular Biology **2010** (17): 666-672.
 10. Cathepsin L Proteolytically Processes Histone H3 During Mouse Embryonic Stem Cell Differentiation, Duncan, E. M., Muratore-Schroeder, T. L., Cook, R. G., Garcia, B. A., Shabanowitz, J., Hunt, D. F. & Allis, C. D., Cell **2008** (135): 284-294.
 11. Signal transduction by receptors with tyrosine kinase activity, Ullrich, A. & Schlessinger, J., Cell **1990** (61): 203-212.
 12. Principles and Techniques of Biochemistry and Molecular Biology, Wilson, K. & Walker, J. **2010**, Cambridge University Press, Cambridge, U.K. 760.
 13. Electrospray interface for liquid chromatographs and mass spectrometers, Whitehouse, C. M., Dreyer, R. N., Yamashita, M. & Fenn, J. B., Analytical Chemistry **1985** (57): 675-679.
 14. Principles and Practice of Biological Mass Spectrometry, Dass, C. **2001**, Desiderio, D. M. & Nibbering, N. M. M., Wiley-Interscience, New York 416.
 15. An Introduction to Quadrupole Ion Trap Mass Spectrometry, March, R. E., Journal of Mass Spectrometry **1997** (32): 351-369.
 16. The quadrupole ion store (quistor) as a novel source for a mass spectrometer, Lawson, G., Journal of Physics E: Scientific Instruments **1973** (6): 357.
 17. Ion Trap Mass Spectrometry, Wong, P., Current Separation **1997**: 85-93.
-

18. Recent improvements in and analytical applications of advanced ion trap technology, Stafford Jr., G. C., Kelley, P. E., Syka, J. E. P., Reynolds, W. E. & Todd, J. F. J., International Journal of Mass Spectrometry and Ion Processes **1984** (60): 85-98.
 19. Fourier transform ion cyclotron resonance mass spectrometry, Marshall, A. G., Accounts of Chemical Research **1985** (18): 316-322.
 20. The Orbitrap: a new mass spectrometer, Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M. & Cooks, G., Journal of Mass Spectrometry **2005** (40): 430-443.
 21. Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer, Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K. & Horning, S., Analytical Chemistry **2006** (78): 2113-2120.
 22. Ion activation methods for tandem mass spectrometry, Sleno, L. & Volmer, D. A., Journal of Mass Spectrometry **2004** (39): 1091-1112.
 23. Wells, J. M. & McLuckey, S. A. (2005) in *Biological Mass Spectrometry*, ed. Burlingame, A. L. Academic Press: 148-185.
 24. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry, Good, D. M., Wirtala, M., McAlister, G. C. & Coon, J. J., Molecular & Cellular Proteomics **2007** (6): 1942-1951.
 25. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F., Proceedings of the National Academy of Sciences of the United States of America **2004** (101): 9528-9533.
 26. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process, Zubarev, R. A., Kelleher, N. L. & McLafferty, F. W., Journal of the American Chemical Society **1998** (120): 3265-3266.
 27. Electron capture versus energetic dissociation of protein ions, Kruger, N. A., Zubarev, R. A., Carpenter, B. K., Kelleher, N. L., Horn, D. M. & McLafferty, F. W., International Journal of Mass Spectrometry **1999** (182-183): 1-5.
 28. Investigation of transferrin structure via novel electron capture dissociation techniques using a hybrid linear ion trap time-of-flight mass spectrometer, Shaw, M. A., Tsuboyama, A., Ostrander, C., Hashimoto, M. & Wantanabe, M., Spectroscopy **2008**.
 29. Analysis of proteins and peptides on a chromatographic timescale by electron-transfer dissociation MS, Udeshi, N. D., Shabanowitz, J., Hunt, D. F. & Rose, K. L., FEBS Journal **2007** (274): 6269-6276.
 30. A Proteomics Grade Electron Transfer Dissociation-Enabled Hybrid Linear Ion Trap-Orbitrap Mass Spectrometer, McAlister, G. C., Berggren, W. T., Griep-Raming, J., Horning, S., Makarov, A., Phanstiel, D., Stafford, G., Swaney, D. L., Syka, J. E. P., Zabrouskov, V. & Coon, J. J., Journal of Proteome Research **2008** (7): 3127-3136.
 31. Front-end Electron Transfer Dissociation, Compton, P. D., **2009**, University of Virginia, Charlottesville, VA, Department of Chemistry.
-

2 Comprehensive Analysis of Phosphorylation Sites in Tensin 1 Reveals Regulation by p38 MAPK

2.1 Abstract

Cell migration involves periodic attachment to and movement along extracellular surfaces. This is achieved through a highly regulated process involving proteins that assemble into complexes to anchor the cell membrane to the extracellular matrix. These structures are called focal adhesions and are imperative for proper migration and promote extracellular signal recognition. Dysregulation of phosphorylation-dependent signal transduction at focal adhesions can lead to the loss of cell polarity, uncontrolled migration and invasion, a dangerous event that causes cancer metastasis. Tensin 1 is a large scaffolding protein that participates in cell migration. Until now, limited tensin 1 phosphorylation has been reported. We present the first comprehensive phosphorylation site analysis of tensin 1.

An isoform-specific association between tensin 1 and protein phosphatase 1 α (PP1 α) in focal adhesions has been observed. PP1 α activity towards tensin 1 in focal adhesions is investigated by determining phosphorylation differences in tensin 1 WT and F302A, a tensin 1 mutant that disrupts PP1 α binding *in vivo*. Characterization of phosphorylation on tensin 1 was completed using tryptic digestion of the entire 1,735 amino acid sequence followed by analysis using high resolution mass spectrometry implementing both collision-activated and electron transfer dissociation methods for PTM site determination. Remarkably, these

methods resulted in 93% sequence coverage of the entire tensin 1 amino acid

sequence and nearly 98% coverage of all phosphorylatable residues (Ser, Thr, Tyr).

The use of immobilized metal affinity chromatography (IMAC) for phosphorylation-specific enrichment with isotopic labeling strategies resulted in the characterization of an incredible 50 tensin 1 Ser/Thr phosphorylation sites without calyculin A-induced PP1 α inhibition, 12 additional sites upon calyculin A treatment and 10 Tyr phosphorylation sites following peroxyvanadate treatment to inhibit tyrosine phosphatases. Adding to the already complex PTM profile of tensin 1, two residues were found to be modified by *O*-linked N-acetylglucosamine (*O*-GlcNAc) as well.

This comprehensive analysis demonstrates that tensin 1 is a highly modified phosphoprotein and contains modifications at sites previously unreported. Consideration of the sequences proximal to modified sites resulted in the discovery that Pro-directed kinases were responsible for a large amount of the total Ser and Thr phosphorylation. Subsequent *in vitro* and *in vivo* studies determined that p38 α MAPK is responsible for a large portion of tensin 1 phosphorylation. These p38 α MAPK-mediated changes in phosphorylation affect binding of several focal adhesion-related proteins including deleted in liver cancer-1 (DLC-1), p130Cas and focal adhesion kinase (FAK). Overall, the data demonstrates that tensin 1 is extensively but specifically phosphorylated by p38 α MAPK and other kinases within focal adhesions. As a result, tensin 1 is now known to be part of a complex PTM-mediated signaling system that includes Ser, Thr and Tyr kinases, PP1 α and RhoGTPases.

2.2 Introduction

2.2.1 Cell Signaling and Protein Phosphorylation

In multicellular organisms, cells are not solitary entities that function without regard to their environment. Instead, they receive extracellular signals in the form of small molecules like growth factors that, upon signal reception, dictate a specific cellular response (1). Receptor proteins on the surface of cellular membranes can recognize unique signaling molecules and transmit the corresponding information to the cell interior via regulated signal propagation through cascades to elicit the proper downstream response (2). In this fashion, cells can respond to many different stimuli including those originating from nearby cells in order to regulate critical processes that promote cell growth, proliferation and cell differentiation (1). Intracellular signal transduction mediated by similar means allows for communication between compartments to induce necessary changes required for proper growth and development. Likewise, cells can communicate from the inside out by releasing signaling molecules in order to elicit responses from other cells.

The ability to transmit signals within cells involves three protein-related changes: 1) PTM of proteins within specialized cellular pathways, 2) altered protein-protein interactions leading to functional changes and 3) fluctuations in protein abundance levels mediated by half lives and protein expression variations (3). The study of protein PTM has garnered serious investigation because post-

translationally modified proteins are known to influence all three signal

transduction modes.

All PTMs are enzymatically regulated and most consist of the addition of small chemical moieties such as phosphate, *O*-GlcNAc, acetyl or methyl groups to amino acid side chains (4). Of all known protein PTMs, phosphorylation is perhaps the most ubiquitous and well studied. This is due in part to the availability several phosphopeptide specific enrichment strategies (e.g. IMAC). Selective IMAC enrichment made possible by chelation with Fe(III) imidodiacetate resin enhances the identification of phosphorylated proteins and peptides using MS (**Figure 2.1**). The ability to exclusively isolate phosphorylated peptides relies on the derivatization of carboxylic acid-containing side chains (Asp, Glu) and C-termini to discourage enrichment based on similar interactions. Enrichment employing this strategy has allowed extensive phosphoproteomic analyses and the identification of phosphorylation targets throughout the proteome. Analyses of this type continue to expand the knowledge of phosphorylation-based signaling networks and remain imperative for gaining a more complete understanding of signaling cascades. Ongoing experiments of this nature are necessary considering that one-third of all human proteins are estimated to be modified with phosphate (5).

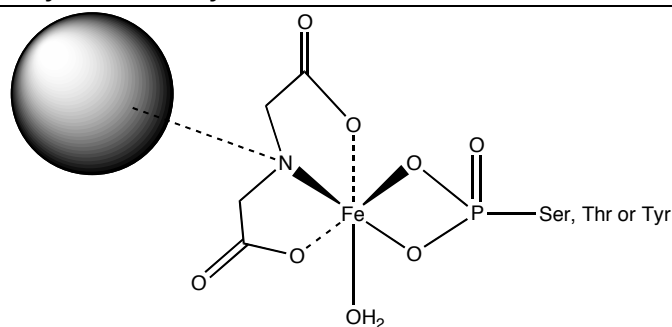


Figure 2.1 Phosphorylated peptide binding to Fe (III) imidodiacetate IMAC POROS® MC 20 resin conjugated to a solid support. Underivatized carboxylic acids interact in a similar fashion with Fe(III) though acid-specific esterification disrupts the interaction. Phosphate release is achieved by Fe(III) reduction to Fe(II) using ascorbic acid or other reducing agents.

Signal transduction by protein phosphorylation is experienced by other macromolecules due to the change in electrostatics induced by the addition of a -2 charge on a neutral Ser, Thr or Tyr side chain at physiological pH (6). This chemical variation can ultimately result in modified protein conformations, unfavorable or increased preferences for protein:protein interactions and changes in cellular location (4). In all cases, modulated functionality can be a consequence.

Conversely, enzymatic *O*-GlcNAcylation of Ser and Thr side chains will instead place an uncharged monosaccharide at the same or a proximal location to prevent phosphate-induced chemical changes (7). The presence of *O*-GlcNAc increases the complexity of Ser and Thr PTM due to complex interplay with nearby or same site phosphate occupancy. *O*-GlcNAc and phosphate PTMs share common attributes that suggest similar biological function. These include similarities in estimated abundance, ubiquitous distribution on proteins of varied function, and a characteristic reversibility associated with PTM addition and removal on relatively equal time scales (8).

The extensive network that directly controls protein phosphorylation consists of nearly 500 kinases, enzymes that catalyze the addition of a single phosphate group from ATP to Ser, Thr and Tyr side chains (5). In the entire human proteome, there is approximately one phosphatase, the enzyme that catalyzes the removal of phosphate from phosphoproteins, in number for every three kinases represented (5). This suggests that phosphatases are more promiscuous than kinases, which generally exhibit very specific substrate selection. The number of proteins involved in these phosphorylation signaling pathways drastically increases when those involved in signal reception and protein:protein interactions used for scaffolding purposes are considered. The ability of these proteins to specifically recognize phosphate addition by kinases and removal by phosphatases provides a regulated system possible for identification of and responses to changes in the local environment (6).

To date, signaling by phosphorylation has been implicated in many critical pathways including those that maintain cellular growth and proliferation, promote cell cycle progression, and regulate gene transcription (2, 6). Intense investigation into phosphorylation's role has stemmed from the discovery that dysregulated phosphorylation is highly prevalent in diseased cells, most often in reference to cancer (9). The consequences of cellular transformation into an uncontrolled cancer state are now obvious though unfortunately, the highly complex mechanisms responsible are not.

2.2.2 Focal Adhesions and Cell Migration

A significant concern regarding some cancers is metastasis, the process by which cancer cells migrate to other areas of an organism and continue to proliferate wildly. Metastasis is facilitated by the ability of cells to become mobile, a process generally referred to as cell migration. However, properly regulated cell migration is by no means a disadvantageous property and serves an integral role in embryogenesis, wound healing, inflammation and other natural processes (10, 11).

Properly controlled cell migration is possible through the formation of thousands of regulated structures that extend through the cell membrane called focal adhesions. These protein complexes, first identified in 1971, work in tandem to anchor cells through the action of α and β heterodimeric, transmembranal integrins that indirectly attach the actin cytoskeleton to the extracellular matrix (ECM) and other surfaces (10-12). Mammals are known to express at least 18 α and 8 β subunits that combine to represent at least 24 different integrin heterodimers that respond differently to various ECM-related stimuli (13). Extracellular tails of integrin heterodimers can bind the ECM and initiate a series of intracellular events that lead to the formation of focal adhesions (11).

Following activation initiated by ECM-binding, focal adhesion protein talin binds an NPxY motif on the β -integrin cytoplasmic tail to disrupt α tail association and produces a conformational change resulting in higher affinity integrin:ECM binding (12, 13). Kindlin proteins function similarly at NxxY motifs to promote

integrin activation (13). As pictured in **Figure 2.2**, cytoplasmic integrin tails bind activated or recruited focal adhesion proteins while extracellular portions remain in contact with the ECM in order to form a more sturdy protein complex (10). The protein component of focal adhesions rapidly changes and is dependent on the intracellular signaling pathway initiated by integrin ligation (11, 13). Regulation of this process involves Ser, Thr and Tyr phosphorylation in order to activate required binding partners (10). PTM-mediated regulation is necessary due to the discovery that certain proteins including FAK and Cas promote motility while others such as vinculin suppress it (11). Once many focal adhesions are formed, cells are anchored to the ECM and allow detachment of rear adhesions to facilitate forward movement. Lamellipodia, or forward extending actin protrusions, promote focal adhesion formation at the leading edge while subsequent detachment at the rear of the migrating cell allows physical movement once the rear of the cell is pulled forward (14). Repeated cycles of this sequence in response to signaling molecules facilitate forward, unidirectional cell migration.

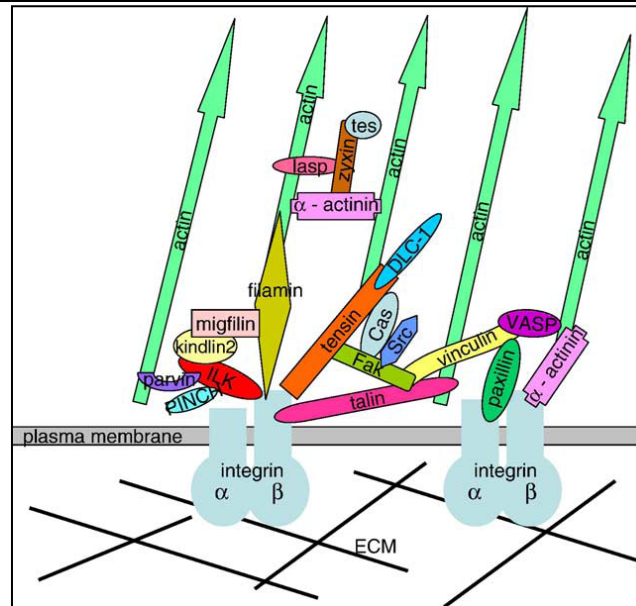


Figure 2.2 Focal adhesions are protein complexes that anchor the cytoskeleton to the extracellular matrix through transmembrane, heterodimeric integrins and many scaffolding proteins. Figure reproduced with permission from (10).

While it is evident that focal adhesions form in order to provide structural anchors to extracellular surfaces, they also serve more generally as receptors to transmit extracellular signals to the cell interior. This can result in changes in proliferation and differentiation of both healthy and cancer cells (15). It is of great importance that focal adhesions are securely regulated due to the realization that kinase dysregulation can cause transformation and because adhesions mediate cell migration for prevention of uncontrolled cell motility and metastasis. Since in-depth details surrounding focal adhesion maturation are scarcely known, study of the process and required regulatory mechanisms could be of utmost importance for disease treatment options (12). The focus of this chapter is an investigation into the regulatory PTMs of a key player in focal adhesion maturation, tensin 1.

2.2.3 Tensin 1

Tensin 1 is a large, cytoplasmic focal adhesion protein that binds cytoskeletal actin filaments and β integrin tails (11, 12, 16, 17). This 1,735 amino acid protein is part of the tensin protein family that includes four distinct proteins with highly conserved extreme N- and C-terminal regions (16). Though the molecular weight of tensin 1 is calculated at 180 kDa, under SDS-PAGE it consistently migrates between 200-250 kDa (16). While tensins 1-3 are similar in both molecular weight and function, cten (C-terminal tensin-like protein or tensin 4) lacks the major actin-binding domain (ABD) located in the N-terminal region of tensins 1-3 (16, 18). Cten is therefore unable to bind actin to focal adhesions, resulting in cytoskeleton:adhesion detachment (18). This property affords a biological switch to cten expression higher than full length tensin 3 in invasive breast cancer cells that directly correlates to cancer metastasis (18). Characterization of cten and tensins 1-3 in focal adhesion structural stability is still being investigated.

With the exception of cten, tensins 1-3 all include structural domains required for cytoskeletal anchoring to integrin tails. Actin binding is accomplished through an N-terminal ABD in tensins 1-3. A C-terminal phosphotyrosine-binding domain (PTB) facilitates attachment to β integrin tails (16). Because structural similarity defines PTBs rather than strict functionality, β integrin attachment at NPXY structural motifs using PTBs actually does not require pTyr (16). The C-terminal region of all four tensins includes an Src-homology 2 (SH2) region through

which tensins bind phosphotyrosine-containing focal adhesion proteins such as p130Cas and focal adhesion kinase (FAK) (16, 19). The distinct locales of tensin 1 binding domains that include two N- and C-terminal focal adhesion-binding domains (FABs) are represented in **Figure 2.3**. An additional ABD is located in the center region of tensin 1 though the central regions in all tensins are structurally dissimilar and otherwise contain no recognizable domains (16).

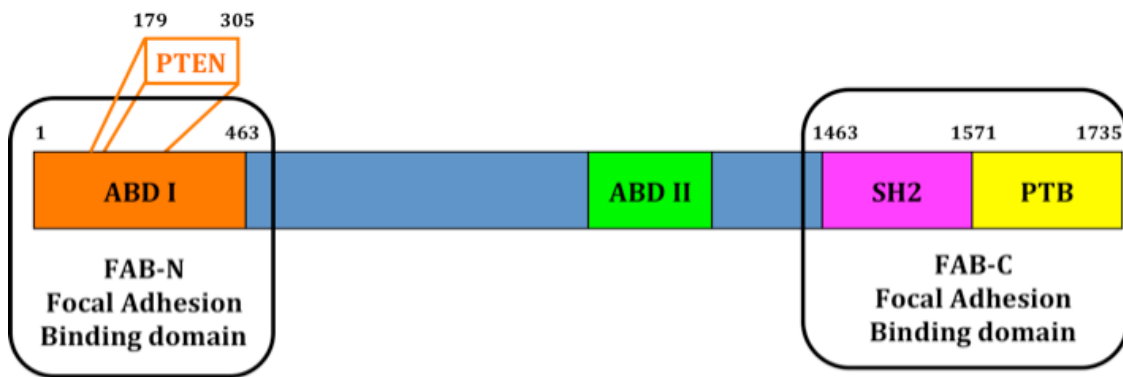


Figure 2.3 Structural domains of human tensin 1. Domains include two actin binding domains (ABDs), a Src homology 2 (SH2) domain for pTyr interactions and a phosphotyrosine binding (PTB) domain that does not require pTyr for protein interactions. Focal adhesion binding domains (FABs) are located at both termini and facilitate interactions with various focal adhesion-related proteins including DLC-1 and PP1 α . An inactive phosphatase and tensin 1 homologue on chromosome 10 (PTEN) domain is located near the N-terminus and facilitates binding to PP1 α . Figure adapted with permission from (16).

The processes that specifically control tensin 1 involvement in focal adhesion assembly are not completely known. Several tensin 1 Ser, Thr and Tyr phosphorylation sites have been identified and reported on PTM database www.phosphosite.org (entry Q9HBL0), though regulation by kinases is still unclear and functional details continue to elude researchers (16). Several interesting observations prompt further investigation. First, tensin 1 expression is drastically downregulated in many cancer cell lines including breast and prostate cancers while

increased cten expression is known to significantly increase (17). Second, tensin 1 cleavage has been implicated in pro-apoptotic pathways involving focal adhesion disruption (16). Third, tensin 1 is found to be actively recruited to focal adhesions after integrin activation (20). This study aims to investigate tensin 1 PTMs that could serve to mediate these important processes.

2.2.4 Protein Phosphatase 1 α

As previously mentioned, phosphorylation of Ser, Thr and Tyr residues are present on focal adhesion proteins and are assumed to contribute to regulation of adhesion assembly. Of course, proper regulation by phosphorylation PTMs requires both kinases and phosphatases, both of which have been already found in focal adhesions. Eto et al. have determined using specific antibodies and co-immunopurification that protein phosphatase 1 α (PP1 α) in fibroblast focal adhesions uniquely co-localizes with tensin 1 with isoform specificity (20). Additionally, it was discovered using immunofluorescence that tensin 1 and PP1 α are not present in originating focal adhesions but are instead localized to maturing focal adhesions away from leading edge protrusions used to initiate cell migration (20). This observation suggests that tensin 1 is actively recruited to maturing focal adhesions following integrin activation. Due to the presence of a colocalized phosphatase, it stands to reason that phosphorylation PTM may play a role in the recruitment process.

The three PP1 isoforms ($\alpha, \beta/\delta, \gamma$), which contain nearly identical catalytic domains and variable C-terminal regions, have been found to specifically bind to other focal adhesion proteins in a similar, isoform specific manner (ie. PP1 δ to FAK) (20). The use of tensin 1 truncations (1-360, 1-740, 1-1462) that co-purified PP1 α showed that the tensin 1 N-terminal PTEN domain was sufficient for PP1 α binding (20). It was also determined that the interaction between tensin 1 and PP1 α occurs at an (R/K)VXF motif characteristic of PP1 α binding to regulatory subunits. Specifically, mutation of all 3 N-terminal tensin 1 VXF motifs to VXA demonstrated that F302 is largely responsible for PP1 α association since the F302A mutation produced an 80% loss of PP1 α (20, 21).

The proposed regulation of tensin 1 by PP1 α prompted Emily Hall, a post-doctoral researcher in David Brautigan's laboratory in the University of Virginia Department of Microbiology, to gauge tensin 1 phosphorylation levels in response to varied phosphatase activity. Treatment of HEK293 cells expressing tagged tensin 1 with calyculin A, a PP1/PP2A phosphatase inhibitor, produces a change in electrophoretic mobility as compared to untreated cells (**Figure 2.4**). Reduced migration under SDS-PAGE conditions is common with increased phosphorylation-induced charge modification. These results suggest that PP1 α may indeed play a role in manipulating phosphorylation PTMs on tensin 1. Treatment with tyrosine phosphatase inhibitor peroxyvanadate produces another shift and suggests that Ser/Thr and Tyr phosphorylation all exist on tensin 1. The focused goal of the current study is to identify specific changes in tensin 1 phosphorylation resulting

from the increased phosphorylation observed in **Figure 2.4** and to develop models that account for these changes.

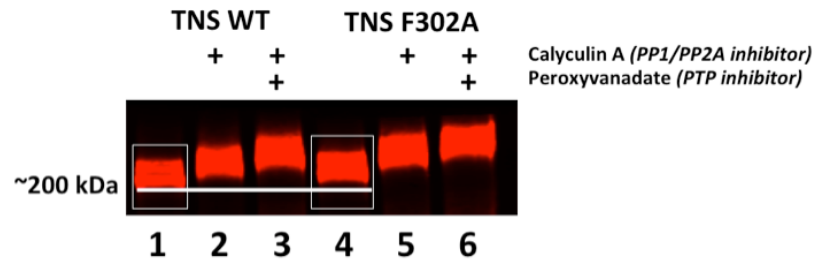


Figure 2.4 Electrophoretic mobility changes under SDS-PAGE for tensin 1 WT and F302A upon phosphatase inhibitor treatment. This reduced mobility suggests modification by phosphorylation.

2.2.5 DLC-1

Deleted in liver cancer-1 (DLC-1) is a tumor suppressing Rho GTPase-activating protein (RhoGAP) that has previously been implicated in cell migration (21). Specifically, Hall and coworkers determined that DLC-1 binds tensin 1 at R1488 because association is significantly reduced with R1488A mutation (21). Interestingly, when tensin 1 is associated with PP1 α it also binds DLC-1, as demonstrated by co-precipitation (21). However, tensin 1 F302A exhibits both reduced PP1 α and DLC-1 binding (21). This suggests that a scaffold consisting of all three proteins can exist only when PP1 α binds tensin 1. The reverse is not true; tensin R1488A did not affect PP1 α binding because this interaction is reserved to the N-terminus.

Observations resulting from DLC-1 bound to tensin 1:PP1 α were surprising. Expression of either tensin 1 mutant, both of which result in loss of DLC-1 binding, exhibited the loss of cell polarity, reduced myosin light chain (MLC)

phosphorylation and lower RhoA activation compared to those expressing tensin 1 WT (21). This is significant because cell migration requires phosphorylation of MLC by MLC kinases to yield cell polarization for proper, unidirectional migration (21). Myosin phosphatase, which regulates MLC phosphorylation, is negatively regulated by a RhoA-activated kinase (21). The described observations clearly demonstrate that disruption of the tensin 1:PP1 α :DLC-1 complex interrupts communication necessary for proper adhesion formation and unidirectional migration. This model is strengthened by the observation that tensin 1 WT expression in tensin 1-null breast cancer cell line MDA MB 231 showed reduced invasion and evidence of regulated migration (21). Taken together, a model can be constructed that brings tensin 1, PP1 α and DLC-1 together to collectively form regulated focal adhesions along with the presence of other proteins. DLC-1 is not assumed to directly influence tensin 1 phosphorylation, but the implications for DLC-1 binding as a result of changes in phosphorylation will be addressed.

2.3 Materials

2.3.1 Reagents

Alltech Associates, Inc. (Deerfield, IL)

Acetyl chloride, anhydrous

D₀-Methanol, anhydrous

D₄-Methanol, anhydrous

Applied Biosystems™ (Carlsbad, CA)

POROS® MC 20 metal chelating packing material, 20 µm diameter

Bio-Rad Laboratories, Inc. (Hercules, CA)

4-15% gradient polyacrylamide gel

Cell Signaling (Danvers, MA)

Anti-p38 MAPK antibody

Anti-phosphorylated p38 MAPK antibody

Anti-phosphorylated Thr-Pro antibody

Mallinckrodt, Inc. (Paris, KY)

Acetonitrile, HPLC grade, ≥99.8% purity

Millipore (Billerica, MA)

Anti-DLC-1 antibody (mouse)

4G10 anti-phosphotyrosine antibody (mouse)

Novagen (Madison, WI)

S-protein agarose

S-protein monoclonal antibody

Open Biosystems, Inc. (Huntsville, AL)

Arrest-In™ Transfection Reagent

Promega (Madison, WI)

Trypsin, sequencing grade (porcine)

PQ Corporation (Valley Forge, PA)

Kasil® 1624

Lithisil®

Roche (Penzberg, Germany)

Chymotrypsin, sequencing grade (bovine)

Sigma Aldrich (St. Louis, MO)

Ammonium bicarbonate

Angiotensin I acetate salt hydrate, ≥ 90% purity (human)

Ascorbic acid

Dithiothreitol

Ethylenediaminetetraacetic acid

Fluoranthene, 99% purity

Formamide

Glacial acetic acid, $\geq 99.9\%$ purity

Iodoacetamide

Iron (III) chloride

Tetramethylammonium silicate

Vasoactive intestinal peptide fragment 1-12, $\geq 97\%$ purity (human, porcine, rat)

Thermo Scientific (Rockford, IL)

Anti-mouse HRP antibody

2.3.2 Equipment & Instrumentation

Agilent Technologies (Palo Alto, CA)

HP1100 high performance liquid chromatograph

Labconco Corporation (Kansas City, MO)

Centrivap centrifugal concentrator

LI-COR® Biosciences (Lincoln, NE)

Odyssey Infrared Imaging System

Polymicro Technologies, LLC (Pheonix, AZ)

360 μm o.d. x 50 μm i.d. polyimide coated fused silica capillary

360 μm o.d. x 75 μm i.d. polyimide coated fused silica capillary

360 μm o.d. x 150 μm i.d. polyimide coated fused silica capillary

Sutter Instrument Co. (Navato, CA)

P-2000 microcapillary laser puller

Thermo Scientific (San Jose, CA/Bremen, Germany)

LTQ-Orbitrap hybrid mass spectrometer

LTQ-FT-ICR hybrid mass spectrometer

YMC Co., Ltd. (Kyoto, Japan)

C18 resin (5-20 μm diameter, 300 \AA pore size)

C18 resin (5 μm diameter, 120 \AA pore size)

Zeus Industrial Products (Orangeburg, SC)

0.060 in o.d. x 0.012 in i.d. Teflon tubing

2.4 Methods

The sample preparation and cell biology-based experiments were completed as described in (22) by Dr. Emily H. Hall, former post-doctoral associate in Dr. David L.

Brautigan's laboratory in the Department of Microbiology at the University of Virginia School of Medicine.

2.4.1 Cell Culture

HEK293 and HEK293T cells were grown in modified Eagle's medium and Dulbecco's modified Eagle's medium, respectively. Both were supplemented with 10% fetal bovine serum. Cultures experienced 5% CO₂ humidified environments and were maintained at 37°C.

2.4.2 Affinity Purification of S-tagged Tensin 1 and Tensin 1 F302A

Full-length S-tagged (pTriEx₄) tensin 1 was mutated at Phe302 to Ala (F302A) using PCR as described previously (21). S-tagged tensin 1 WT and F302A were transiently transfected into separate HEK 293T cell cultures using Arrest-In™ according to manufacturer's instruction. Five 15 cm plates of cells expressing each tensin 1 construct were grown to 80% confluency as described in (22). Control cultures were generated using DMSO treatment while additional cultures were treated with 25 nM calyculin A for 30 min in order to preserve *in vivo* Ser/Thr phosphorylation. Cells were then collected using centrifugation at 1500 x *g* for 5 min. To preserve *in vivo* Tyr phosphorylation, cells were treated with 1 mM peroxyvanadate for 30 min. Following detergent-mediated lysis, cellular extract was incubated with 60 μL of S-protein agarose for 3 h. Agarose beads with bound tensin 1 WT or F302A were later washed a total of five times with varying salt concentrations to remove nonspecifically bound proteins prior to digestion (22).

2.4.3 Immunoblot Analysis of S-tagged Tensin 1

HEK293 cell cultures transfected with either S-tagged tensin 1 WT or F302A were each treated with DMSO (control) or calyculin A as previously described, resulting in 4 experimental conditions: (1) tensin 1 WT + DMSO, (2) tensin 1 WT + calyculin A, (3) tensin 1 F302A + DMSO and (4) tensin 1 F302A + calyculin A. Post-lysis, whole cell extracts were resolved on a 4-15% gradient criterion Tris-HCl polyacrylamide gel after which proteins were transferred to a nitrocellulose membrane and immunoblotted against S-tag (tensin 1), phosphorylated Thr-Pro or Tyr residues (4G10) (22). Immunoblots were then scanned by the Odyssey Infrared Imagine System and analyzed using Odyssey Application Software version 2.0.41.

2.4.4 Proteolytic Digestions and Sample Preparation for Mass Analysis

In silico digestions were performed on S-tagged tensin 1 WT and F302A sequences (N-terminally S-tagged UniProt Accession B2RU35) using ExPASy PeptideMass digest simulator at <http://expasy.org/tools/peptide-mass.html>.

Affinity purified S-tagged tensin 1 from each of the four experimental conditions was retained on agarose beads and washed three times using 100 μ L aliquots of 100 mM ammonium bicarbonate, pH 7.9. While still retained on S-protein agarose at room temperature, tensin 1 WT and F302A Cys residues were reduced for 1.5 h using 10 μ L of 2.16 mM dithiothreitol in 100 mM ammonium bicarbonate and carbamidomethylated in the dark for 45 min using 10 μ L of 4.32 mM iodoacetamide in 100 mM ammonium bicarbonate. On beads digestion of

tensin 1 proceeded on a shaker at room temperature for 7 h using 0.75 μg trypsin in 100 mM ammonium bicarbonate. A secondary endoproteinase chymotrypsin digestion was performed in order to generate smaller peptides for the determination of phosphorylation sites on tensin 1 tryptic peptide Asp1199 to Arg1248. Subdigestion conditions were similar to those previously mentioned and used 75 ng chymotrypsin in 100 mM ammonium bicarbonate, pH 7.9. All digestions were quenched using 2-5 μL of glacial acetic acid for storage at -35°C until mass analysis.

2.4.5 HPLC Capillary Column Assembly

An 8 cm precolumn was constructed using 360 μm o.d. x 75 μm i.d. fused silica capillary with a 2 mm Kasil® 1624 frit and packed with 5 cm of irregular C18 reverse phase resin (5-20 μm diameter, 300 Å pore size). A 9 cm analytical column was constructed using 360 μm o.d. x 50 μm i.d. fused silica capillary containing a 1 mm Lithisil® frit. The analytical column was packed with 7 cm of regular C18 resin (5 μm diameter, 120 Å pore size) and equipped with a laser-pulled, integrated electrospray emitter tip as previously described (23). Both columns were prepped using several picomoles of angiotensin in order to optimize chromatographic performance prior to use for sample analysis.

2.4.6 HPLC-MS/MS Analysis

For mass spectrometric analysis, typically a 5% fraction of the acidified protein digest was pressure loaded onto the precolumn at a flow rate of 1 $\mu\text{L}/\text{min}$. After sample and 250 fmol internal standard (angiotensin and vasoactive intestinal peptide) loading, a desalting rinse of the precolumn using 0.1 M acetic acid in water was completed for approximately 10 min at similar flow rates. The precolumn was then butt connected to the analytical column using a 1 cm Teflon sleeve (0.060 in o.d. x 0.012 in i.d). Tryptic peptides were gradient eluted at a split flow-generated rate of 60 nL/min from the HPLC column into the mass spectrometer using an Agilent 1100 chromatograph and linear LC gradient as follows: 0-60% Solvent B in 60 min, 60-100% Solvent B in 5 min, 100% Solvent B hold for 5 min, 100% - 0% B in 5 min, 100% Solvent A for 20 min (Solvent A: 0.1 M acetic acid in water and Solvent B: 70% acetonitrile in 0.1 M acetic acid in water).

Mass spectrometric analyses incorporated an initial high resolution (60,000 at m/z 400) MS1 survey scan in the Orbitrap or FT-ICR cell, occurring simultaneously with the acquisition of 8 to 10 low resolution, data dependent MS/MS scans measured in the linear ion trap. Both CAD and front-end ETD analyses were completed to yield complementary data sets for optimal peptide characterization. Data dependence parameters for all MS/MS data acquisition included monoisotopic precursor selection "on," a repeat count of 2, repeat duration of 30 s, exclusion list duration of 20 s, and precursor isolation window of 3 m/z . ETD parameters included 30 - 75 ms reaction time, charge state rejection "on" for z

= +1 and +2, FTMS AGC target of 2e5, ITMSⁿ AGC target of 1e4, and azulene ETD reagent target of 2e5.

2.4.7 Phosphorylation Analysis using Differential Isotopic Labeling and Immobilized Metal Affinity Chromatography for Phosphopeptide Enrichment

In order to prevent enrichment of peptides containing carboxylic acids, tensin 1 WT and F302A tryptic peptides were esterified on Asp, Glu and C-terminal residues using D₀-methanolic HCl and D₃-methanolic HCl, respectively. Esterification reagents were prepared with 160 µL anhydrous acetyl chloride added dropwise to 1 mL anhydrous D₀- or D₄-methanol. The reaction mixture was stirred for 5 min after which 40 µL were added to each peptide sample. The reaction was completed twice to ensure complete conversion of acidic residues. Following esterification and mass analysis of a fraction of each digest to ensure conversion and estimate peptide abundances, tensin 1 WT and F302A peptides were combined in equal amounts and enriched using IMAC. Two standard peptides were added to the esterified sample in order to gauge IMAC enrichment efficiency. Each standard peptide, STsLVEGR and DRVyIHPF (angio II phosphate), was added in final concentrations of 250 fmol.

A 360 µm o.d. x 150 µm i.d. fused silica capillary IMAC column with a 2 cm Kasil® 1624 frit was packed with 6 cm of POROS® MC 20 imidodiacetate packing material. The column was cleaned using several pressure-mediated solvent exchanges as follows: a Millipore® water rinse for 5 min at 20 µL/min, 50 mM

EDTA (pH 9) wash for 10 min at 20 $\mu\text{L}/\text{min}$, and Millipore® water rinse for 5 min at 20 $\mu\text{L}/\text{min}$. The column was activated for phosphopeptide binding using 100 mM FeCl_3 for 10 min at 20 $\mu\text{L}/\text{min}$. A rinse using 0.01% acetic acid in water at 1 $\mu\text{L}/\text{min}$ for 25 min was completed prior to sample loading in order to equilibrate the column pH. The esterified peptide sample was reconstituted in 1:1:1 solution (methanol: acetonitrile: 0.01% acetic acid) and loaded onto the IMAC column at a flow rate of 1 $\mu\text{L}/\text{min}$. Following sample loading, the column was rinsed with 25 μL of 1:1:1 and 15 μL of 0.01% acetic acid in water prior to phosphopeptide elution onto a Teflon sleeve-connected precolumn using 10 μL 250 mM ascorbic acid (pH 2). The IMAC-precolumn conjugate was rinsed with 5 μL of 0.1% acetic acid in water to assure removal of ascorbic acid from the precolumn prior to mass analysis. After column separation, the precolumn was washed alone with 0.1 M acetic acid for 10 min prior to loading 250 fmol angiotensin (angio) and vasoactive intestinal peptide (vaso) for relative quantitation purposes. Reattachment with the analytical column and gradient elution into the mass spectrometer proceeded as previously described.

Mass analysis parameters for the phosphopeptide enrichment sample were similar to methods previously outlined. CAD was used exclusively for tandem MS analysis. Ten CAD MS/MS spectra were acquired data-dependently following each high resolution MS1 scan without using charge state rejection during precursor ion selection.

2.4.8 Mass Spectrometric Data Analysis

For all mass analyses, MS/MS peak lists were generated using Bioworks Browser (version 3.3.1 SP1) and the data were searched against a human tensin 1 WT and F302A protein sequence database (gi 187954597, Accession BC140942.1, NCBI) using the Open Mass Spectrometry Search Algorithm (OMSSA, version 2.1.1). Searches of CAD and ETD data shared the following parameters: trypsin enzyme specificity, E-value cutoff of "1," missed cleavage count of 3, variable modifications carbamidomethylation of Cys, oxidation of Met, phosphorylation of Ser, Thr and Tyr, *O*-GlcNAcylation of Ser, Thr, and ± 0.01 Da and ± 0.35 Da for precursor and product ion masses, respectively. OMSSA provided the removal of reduced charge species within ETD data prior to searches. While OMSSA results were used to guide data analysis, all pertinent spectra were confirmed by manual interpretation of the MS/MS spectra to ensure confident peptide and PTM identification.

2.4.9 *In Vitro* Kinase Assays

S-tagged tensin 1 WT and F302A were purified from HEK293 cells as previously described using S-protein agarose. Beads were incubated in either kinase reaction buffer alone (50 mM Tris-HCl, pH 7.5, 0.1% mercaptoethanol, 0.2 mM Na₃VO₄, and 1 mM DTT), kinase buffer with added 10 μ M kinase inhibitors (GSK β 3 inhibitor II, SB220025, or staurosporine) or with 100 ng of p38 α MAPK purified kinase. For each experiment, radioactive ³²P-ATP was added in a Mg/Mn buffer and incubated with purified tensin 1 and kinase for various time periods at 30°C and resolved by SDS-PAGE as described in (22).

2.4.10 DLC-1 Binding Assays

S-protein agarose purified S-tagged tensin 1 was incubated with \pm p38 α MAPK with cold ATP. Purified DLC-1 was incubated with S-tagged tensin \pm p38 α MAPK incubation while retained on beads for 2 h. Following incubation, the S-protein beads were washed, boiled and the corresponding proteins resolved using SDS-PAGE. Proteins were then transferred to a nitrocellulose gel for immunoblot analysis using anti-DLC-1 and anti-S-tag to determine DLC-1 interactions as described in (22).

2.4.11 Hyperosmotic Sorbitol Cell Treatment

HEK293 cells either transfected with S-tagged tensin 1 or reagents alone (control) were treated with 0.4 M sorbitol for 20 min to stimulate p38 α MAPK activity and lysed. S-protein agarose was used to purify S-tagged tensin 1 and any interacting proteins. DLC-1 binding was probed as previously mentioned. In addition, proteins were eluted from agarose, resolved using SDS-PAGE and transferred to nitrocellulose membranes to immunoblot for the following conditions and proteins: anti-pThr-Pro, anti-DLC-1, anti-S-tag (tensin 1), anti-pTyr, anti-p130Cas, and anti-focal adhesion kinase (FAK). In addition, extracts from cells \pm sorbitol treatment were immunoblotted using anti-p38 MAPK and anti-phosphorylated p38 MAPK.

2.5 Results & Discussion

2.5.1 S-tagged Tensin 1 Purification

HEK293 cells were transfected with plasmids encoding either tensin 1 WT (**Figure 2.5**) or tensin 1 F302A. The protein product of S-tagged tensin 1 F302A contains an identical sequence disregarding the single site mutation. Purification of both proteins resulted from the high affinity interaction between the N-terminal, 15 amino acid S-peptide tag on tensin 1 and 103 residue S-protein on agarose.

S-tagged-TNS1: KETAAAKFERQHMS...LEF

```

1 MSVSRTMEDS CELDLVYVTE RIIAVSFPST ANEENFRSNL REVAQMLKSK HGGNYLLFNL
61 SERRPDITKL HAKVLEFGWP DLHTPALEKI CSICKAMDTW LNADPHNVVV LHNKGNRGRI
121 GVVIAAYMHY SNISASADQA LDRFAMKRFY EDKIVPIGQP SQRRYVHYFS GLLSGSIKMN
181 NKPLFLHHVI MHGIPNFESK GGCRPFLRIY QAMQPVYTSG IYNI PGDSQT SVCITIEPGL
241 LLKGDILLKC YHKKFRSPAR DVIFRVQFHT CAIHDLGVVF GKEDLDDAFK DDRFPEYGKV
301 EFVFSYGPEK IQGMEHLENG PSVSVDYNTS DPLIRWDSYD NFSGHRDDGM EEVVGHTQGP
361 LDGSLYAKVK KKDSLHGSTG AVNATRPTLS ATPNHVEHTL SVSSDSGNST ASTKTDKTDE
421 PVPGASSATA ALSPQEKREL DRLLSGFGL EREKQGAMYHT QHLRSRPAGG SAVPSSGRHV
481 VPAQVHVNGG ALASERETDI LDDELPNQDG HSAGSMGTLS SLDGVTNTSE GGYPEALSPL
541 TNGLDKSYPM EPMVNGGGYP YESASRAGPA HAGHTAPMRP SYSAQEGLAG YQREGPHPAW
601 PQPVTTSHYA HDPGGMFRSQ SFSEAEQPLP PAPVRRGSSR EAVQRGLNSW QQQQQQQQQP
661 RPPPRQQERA HLESLVASRP SPQPLAETPI PSLPEFPRAA SQQEIEQSIE TLNMLMLDLE
721 PASAAAPLHK QSVPGAWPG ASPLSSQPLS GSSRQSHPLT QSRSGYIPSG HSLGTPEPAP
781 RASLESVPPG RSYSPDYQP CLAGPNQDFH SKSPASSSLP AFLPTTHSPP GPQQPPASLP
841 GLTAQPLLSP KEATSDPSRT PEEEEPLNEG LVAAHRVAGVQ AREKQPAEPP APLRRRAASD
901 GQYENQSPEA TSPRSPGVRV PVQCVSPELA LTIALNPGGR PKEPHLHSYK EAFEEMEGTS
961 PSSPPPSGVR SPPGLAKTPL SALGLKPHNP ADILLHPTGV TRRRIQPEED EGKVVVRLSE
1021 EPRSYVESVA RTAVAGPRAQ DSEPKSFSAP ATQAYGHEIP LRNGTLGGSF VSPSPLSTSS
1081 PILSADSTSV GSFPSGESSD QGPRTPQTPL LESGFRSGSL GQPSPSAQRN YQSSSPLPTV
1141 GSSYSSPDYS LQHFSSSPES QARAQFSVAG VHTVPGSPQA RHRTVGTNTF PSPGFWRAI
1201 NPSMAAPSSP SLSHHQMMGP PGTGFHGSTV SSPQSSAATT PGSPSLCRHP AGVYQVSGLH
1261 NKVATTPGSP SLGRHPGAHQ GNLAGLHNS AIASPGSPSL GRHLGGSGSV VPGSPCLDRH
1321 VAYGGYSTPE DRRPTLSRQS SASGYQAPST PSFPVSPAYY PGLSSPATSP SPDSAAFRQG
1381 SPTPALPEKR RMSVGDRAGS LPNYATINGK VSSPVASGMS SPSSGSTVSF SHTLPDFSKY
1441 SMPDNPETR AKVKFVQDTS KYWYKPEISR EQAIALLDQ EPGAFIIRDS HSFRGAYGLA
1501 MKVSSPPPTI MQQNKKGDMT HELVRHFLIE TGPRGVKLGK CPNEPNFGSL SALVYQHSII
1561 PLALPCKLVI PNRDPTDESK DSSGPANSTA DLLKQGAACN VLFVNSVDME SLTGPOAISK
1621 ATSETLAADP TPAATIVHFK VSAQGITLTD NQRKLFRRH YPLNTVTFCD LDPQERKWMK
1681 TEGGAPAKLF GFVARKQGST TDNACHLFAE LDPNQPASAI VNFVSKVMLN AGQKR

```

Figure 2.5 Full length human tensin 1 WT amino acid sequence. Sequence above the full length tensin 1 sequence corresponds to amino acids contributed by the S-tag and extended linker region. Sequence is numbered based on untagged tensin 1 WT.

Epitope protein tags provide the opportunity to selectively purify proteins of interest. Commonly used epitopes consist of whole proteins or small peptides. Initially, a green fluorescent protein (GFP)-tagged tensin 1 construct was used. However, due to a large degree of contaminating proteins that reproducibly were pulled down using the GFP system, it was discarded for the unexpectedly clean S-tag pulldown. The S-protein affinity tag exploits the high affinity binding characteristics of S-protein and S-peptide, two products generated by subtilisin cleavage of RNaseA (24). **Figure 2.6** displays purification of the S-peptide-tagged tensin 1 F302A with calyculin A treatment as visualized using SDS-PAGE and Western blot analysis.

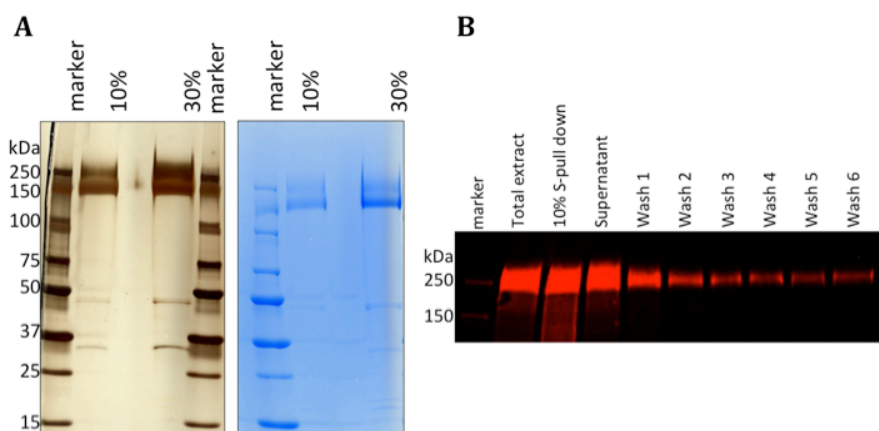


Figure 2.6 Purification of tensin 1 F302A with calyculin A treatment from HEK293T cells. Purifications were visualized by (A) SDS-PAGE analysis with silver stain and GelCode® staining and (B) Western blot analysis using anti-S-tag antibody. Silver and GelCode® stained gel separations of sample fractions suggest a highly pure sample of tensin 1 (~250 kDa) with very few contaminating proteins.

2.5.2 *In silico* Proteolytic Digestion

An ideal proteomics experiment allows for the analysis of intact proteins in order to gain understanding of PTMs that exist simultaneously on protein macromolecules in order to result in unique biological outcomes. However, efficient

methods are not yet available for the study of intact high molecular weight proteins using current mass spectrometry, the method of choice for PTM identification.

Therefore, proteolytic digestion of proteins to peptides amenable for mass analysis is necessary. Proteins of high molecular weight (>50 kDa) in particular present a challenge to proteomic studies not only because increased sample complexity results upon digestion, but also that it is unlikely that a single endoproteinase will generate peptides easily analyzed by mass spectrometry which collectively represent the entire protein sequence. Problematic proteolytic peptides of extreme size fall into one of two categories: (1) small, hydrophilic peptides that may not be retained by commonly used reverse phase chromatographical methods, or (2) large, hydrophobic peptides that are retained indefinitely without regard to gradient elution. Sample loss resulting from either situation can be avoided using optimal proteolytic digestion conditions; this is achieved by exploiting more than one endoproteinases to maximize total protein sequence coverage.

Whereas the average protein molecular weight is close to 50 kDa, the molecular weight of tensin 1 is uncommonly high (>185 kDa) and includes a total of 362 phosphorylatable Ser, Thr and Tyr residues. *In silico* digestions afford mass spectrometrists the ability to compare proteolytic digestions by any endoproteinase in order to choose an enzyme that provides optimum peptide size and hydrophobicity throughout the entire protein sequence. Comparison of tensin 1 peptides generated by commonly used endoproteinases resulted in the choice of trypsin, which specifically cleaves amide bonds C-terminal to basic amino acids Lys

and Arg. Tryptic digestions are normally reserved for CAD analyses exclusively due to the production of short peptides with $z=+2$. However, due to the desirable length (8 – 20 residues) of tensin 1 tryptic peptides throughout the entire sequence and the likelihood that some peptides will exhibit $z>+2$ based on internal basic residues (His and/or Lys-Pro, Arg-Pro motifs) and backbone amides, it was considered optimum for tensin 1 analysis by CAD and ETD MS (**Table 2.1**). Digestions generated by endoproteinases Asp-N, Glu-C, Arg-C and chymotrypsin resulted in a large number of peptides unfavorable for reverse phase HPLC analysis (not shown).

Mass (Da)	Position	Peptide Sequence	# STY	Comments
579.2919	1-5	LEFMSVSR	2	+
1902.8459	6-21	TMEDSCELDLVYVTER	4	+
1794.9020	22-37	IIAVSFPSTANEENFR	3	+
489.2780	38-41	SNLR	1	?
818.4440	42-48	EVAQMLK	0	+
234.1448	49-50	SK	1	-
1519.7652	51-63	HGGNYLLFNLSER	2	+
729.4253	64-69	RPDITK	1	?
468.2929	70-73	LHAK	0	?
1851.9639	74-89	VLEFGWPDLHTPALEK	1	+
666.3313	90-95	ICSICK	1	?
2174.0811	96-114	AMDTWLNADPHNVVVLHMK	1	+
346.1833	115-117	GNR	0	-
232.1404	118-119	GR	0	-
2565.2765	120-143	IGVVIAAYMHYSNISASADQ	5	+
496.2588	144-147	FAMK	0	?
175.1189	148	R	0	-
701.3141	149-153	FYEDK	1	?
1094.6316	154-163	IVPIGQPSQR	1	+
175.1189	164	R	0	-
1570.8264	165-178	YVHYFSGLLSGSIK	5	+
2603.3373	179-200	MNNKPLFLHHVIMHGIPNFESK	1	+

905.4774	201-208	GGCRPFLR	0	+
3812.9489	209-243	IYQAMQPVYTSGIYNIPGDSQTSVCITIEP GLLLK	9	+
658.4134	244-249	GDILLK	0	+
550.2442	250-253	CYHK	1	?
147.1128	254	K	0	-
322.1873	255-256	FR	0	-
430.2408	257-260	SPAR	1	?
649.3668	261-265	DVIFR	0	+
1870.9632	266-282	VQFHTCAIHDLGVVFGK	1	+
952.4258	283-290	EDLDDAFK	0	+
405.1728	291-293	DDR	0	-
740.3613	294-299	FPEY GK	1	+
1301.6412	300-310	VEFVFSYGPEK	2	+
2771.3304	311-335	IQGMEHLENGPSVSDYNTSDPLIR	5	+
1383.5712	336-346	WDSYDNFSGHR	3	+
2318.0605	347-368	DDGMEEVVGHTQGPLDGSLYAK	3	+
246.1812	369-370	VK	0	-
147.1128	371	K	0	-
147.1128	372	K	0	-
4192.0137	373-414	DSLHGSTGAVNATRPTLSATPNHVE HTLSVSSDSGNSTASTK	16	+
363.1874	415-417	TDK	1	-
1955.9556	418-437	TDEPVP GASSATAALSPQEK	5	+
175.1189	438	R	0	-
532.2725	439-442	ELDR	0	?
991.5571	443-451	LLSGFLER	1	+
276.1554	452-453	EK	0	-
1341.6480	454-464	QGAMYHTQH LR	2	+
1285.6607	465-478	SRPAGGS AVPSSGR	4	+
1840.9776	479-496	HVVPAQVHVNGGALASER	1	+
5118.3491	497-546	ETDILDEL PNQDGHSAGSMGTLSSLD GVTNTSEGGY PEALSPLTNGLDK	12	+
2191.9423	547-566	SYPM EPMVNGGGYPYESASR	6	+
2781.3273	567-593	AGPAHAGHTAPMRPSYSAQEGLAGYQR	5	+
2802.2841	594-618	EGPHPAWPQPVT TSHYAHDPSGMFR	5	+
1839.9235	619-635	SQSFSEAEPQLPPAPVR	3	+
463.2259	636-640	GGSSR	2	?
602.3256	641-645	EAVQR	0	?

2429.2181	646-665	GLNSWQQQQQQQQPRPPPR	1	+
560.2787	666-669	QQER	0	-
3136.6789	670-698	AHLESLVASRSPQPLAETPIPSLPEFPR	5	+
3449.7290	699-730	AASQQEIEQSIETLNMLMLDLEPASA AAPLHK	4	+
2340.1578	731-754	SQSVPGAWPGASPLSSQPLSGSSR	8	+
1053.5436	755-763	QSHPLTQSR	3	+
1822.9082	764-781	SGYIPSGHSLGTPEPAPR	5	+
1012.5421	782-791	ASLESVPPGR	2	+
2417.0503	792-812	SYSPYDYQPCLAGPNQDFHSK	6	+
3872.0592	813-851	SPASSSLPAFLPTTHSPPGPQQPPASL PGLTAQPLLSPK	10	+
862.3901	852-859	EATSDPSR	3	+
1803.9235	860-875	TPEEEPLNLEGLVAHR	1	+
700.4100	876-882	VAGVQAR	0	+
276.1554	883-884	EK	0	-
1075.5894	885-894	QPAEPPAPLR	0	+
175.1189	895	R	0	-
175.1189	896	R	0	-
1907.8366	897-914	AASDGQYENQSPEATSPR	5	+
515.2936	915-919	SPGVR	1	?
2347.2802	920-942	SPVQCVSPELALTIALNPGGRPK	3	+
1010.5054	943-950	EPHLHSYK	2	+
2090.9335	951-970	EAFEEMEGTSPSSPPPSGVR	5	+
669.3930	971-977	SPPGLAK	1	+
2618.4776	978-1002	TPLSALGLKPHNPADILLHPTGVTR	4	+
175.1189	1003	R	0	-
175.1189	1004	R	0	-
1044.4844	1005-1013	IQPEEDEGK	0	+
472.3242	1014-1017	VVVR	0	?
730.3730	1018-1023	LSEEPR	1	+
910.4628	1024-1031	SYVESVAR	3	+
671.3835	1032-1038	TAVAGPR	1	+
774.3628	1039-1045	AQDSEPK	1	+
1844.9289	1046-1062	SFSAPATQAYGHEIPLR	4	+
4053.9159	1063-1104	NGTLGGSFVSPSPLSTSSPILSADSTSV GSFSPGESSDQGPR	16	+
1345.7110	1105-1116	TPTQPLLESGFR	3	+
1271.6338	1117-1129	SGSLGQPSPSAQR	4	+

3677.6626	1130-1163	NYQSSSPLPTVGSSYSSPDYSLQHFSSS PESQAR	16	+
1808.9402	1164-1181	AQFSVAGVHTVPGSPQAR	3	+
312.1778	1182-1183	HR	0	-
1387.6964	1184-1197	TVGTNTPPSPGFGR	4	+
175.1189	1198	R	0	-
4903.2654	1199-1248	AINPSMAAPSSPSLSHHQMMGPPGTG FHGSTVSSPQSSAATTPGSPSLCR	16	+
1506.7811	1249-1262	HPAGVYQVSGLHMK	2	+
1142.6164	1263-1274	VATTPGSPSLGR	4	+
2690.3505	1275-1302	HPGAHQGNLASGLHSNAIASPGSPSLGR	5	+
1637.8064	1303-1319	HLGGSGSVVPGSPCLDR	3	+
1451.6549	1320-1332	HVAYGGYSTPEDR	4	+
729.4366	1333-1338	RPTLSR	2	+
4017.8777	1339-1378	QSSASGYQAPSTPSFPVSPAYYPGLS SPATSPSPDSAAFR	16	+
1124.5946	1379-1389	QGSPTPALPEK	2	+
175.1189	1390	R	0	-
175.1189	1391	R	0	-
664.3083	1392-1397	MSVGDR	1	+
1305.6797	1398-1410	AGSLPNYATINGK	3	+
2840.3407	1411-1439	VSSPVASGMSSPSGGSTVSFSHTLPDFSK	12	+
1296.5525	1440-1450	YSMPDNSPETR	4	+
218.1499	1451-1452	AK	0	-
246.1812	1453-1454	VK	0	-
824.4148	1455-1461	FVQDTSK	2	+
1241.6313	1462-1470	YWYKPEISR	3	+
885.5404	1471-1478	EQAIALLK	0	+
1145.5949	1479-1488	DQEPGAFIIR	0	+
748.3373	1489-1494	DSHSFR	2	+
810.4178	1495-1502	GAYGLAMK	1	+
1426.7359	1503-1515	VSSPPPTIMQQNK	3	+
147.1128	1516	K	0	-
1057.5095	1517-1525	GDMTHELVR	1	+
1069.5789	1526-1534	HFLIETGPR	1	+
303.2027	1535-1537	GVK	0	-
260.1968	1538-1539	LK	0	-
2968.5059	1540-1567	GCPNEPNFGSLVYQHSIPLALPCK	4	+
711.4512	1568-1573	LVIPNR	0	+

791.3417	1574-1580	DPTDESK	2	+
1375.6699	1581-1594	DSSGPANSTADLLK	4	+
2693.3273	1595-1620	QGAACNVLFINSVDMESLTGPQAISK	4	+
2041.0600	1621-1640	ATSETLAADPTPAATIVHFK	5	+
1402.7285	1641-1653	VSAQGITLTDNQR	3	+
147.1128	1654	K	0	-
582.3398	1655-1658	LFFR	0	?
175.1189	1659	R	0	-
2047.9542	1660-1676	HYPLNTVTFCDLDPQER	3	+
147.1128	1677	K	0	-
464.2326	1678-1680	WMK	0	-
730.3730	1681-1688	TEGGAPAK	1	+
809.4668	1689-1695	LFGFVAR	0	+
147.1128	1696	K	0	-
3174.5160	1697-1726	QGSTTDNACHLFAELDPNQPASAIVN FVSK	5	+
860.4658	1727-1734	VMLNAGQK	0	+
175.1189	1735	R	0	-

Table 2.1 *In silico* tryptic digestion of full length tensin 1 WT. Bolded letters correspond to those belonging to the N-terminal S-tag. Predicted unmodified masses, residue positions and number of phosphorylatable residues (STY) per peptide are listed. Peptides expected to be detected are denoted with a "+" under the Comments header, questionable peptides denoted by a "?", those not expected to be detected by a "-". This suggests the vast majority of all phosphorylatable residues are included in peptides that should be easily detected.

2.5.3 LC-MS/MS Analyses of Tensin 1 WT & F302A ± calyculin A treatment

2.5.3.1 Base Peak Chromatogram Analysis Illustrates Sample Purity

Mass spectrometric analyses were conducted on tensin 1 WT and F302A under untreated or calyculin A treatment. Tryptic digests corresponding to each condition were each analyzed using CAD and front-end ETD to yield complementary data sets on either an LTQ-Orbitrap or LTQ-FT-ICR high resolution mass spectrometer. In order to adequately separate the complex mixture of tryptic peptides for each analysis, all tensin 1 digests were subjected to a 90 min HPLC

gradient. From this analysis, a base peak chromatogram was generated. Base peak chromatograms display elution profiles of the most abundant ionic species for the duration of the gradient using MS1 data and illustrate the degree of sample complexity. The total ion chromatogram (TIC), on the other hand, integrates ion current from all coeluting ionic species. **Figure 2.7** shows base peak analysis of tryptic peptides from tensin 1 F302A with calyculin A treatment. From this chromatogram it is clear that the tryptic digest is efficiently separated for analysis. Due to variable hydrophobicity, these peptides are uniquely retained and selectively eluted over the entire HPLC gradient. The ability to successfully separate the digest over the gradient time period results in the opportunity for the most informative MS2 analysis.

Ideal protein purifications isolate the protein of interest exclusively. However, due to non-specific protein interactions with the affinity material, contaminating proteins are common. As previously mentioned, a GFP-tensin 1 construct was initially used to purify tensin 1 using an anti-GFP antibody. Contaminating proteins of varied abundances prevented the complete analysis of the entire tensin 1 sequence due to competition for data dependent mass analysis (not shown). The use of the S-tag for purification of tensin 1 resulted in highly pure protein that promoted successful analysis by MS. Peptides labeled within the base peak chromatogram display that the most abundant protein in the analysis is tensin 1. Incredibly, virtually every peak in the base peak chromatogram corresponds to a tensin 1 tryptic peptide and demonstrates an exceptionally clean purification.

As evident in the labeled chromatogram, peptide abundances and chromatographical peak shapes can vary within a single analysis. This is due to slight differences in ionization, separation and digestion efficiencies. Peptide elution patterns, abundances and the high degree of sample complexity represented in **Figure 2.7** are typical of all tryptic tensin 1 mass analyses (data not shown). However, two major exceptions to this pattern exist and result in differences between the four tensin 1 conditions: post-translational modification abundances and peptides that include the F302A site mutation.

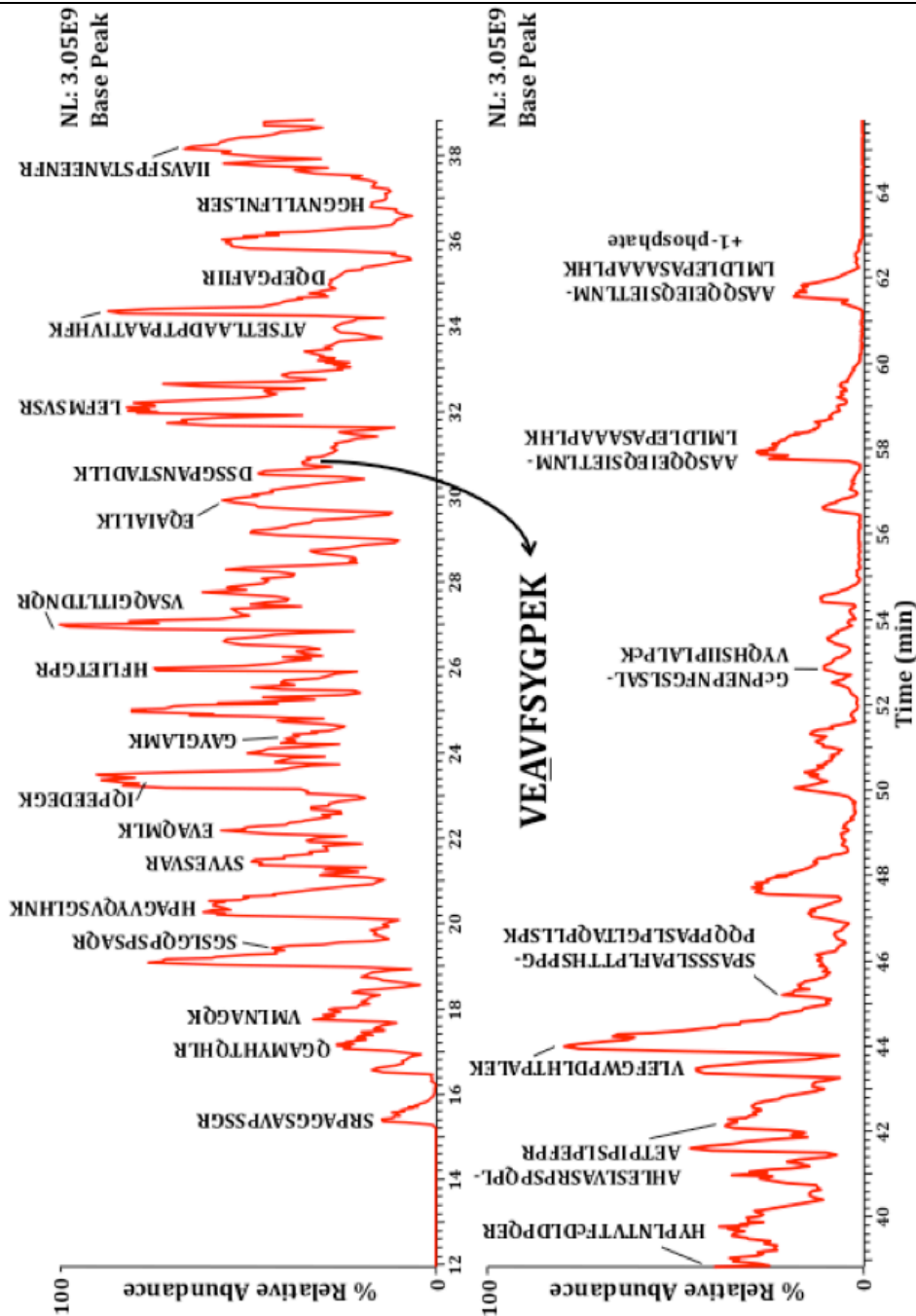


Figure 2.7 Base peak chromatogram resulting from LC-MS/MS analysis of S-tagged tensin 1 F302A with calyculin A treatment. Several tensin 1 peptides are labeled and demonstrate that the sample purity and proteolytic digestion efficiency resulted in highly abundant, tryptic tensin 1 peptides including phosphorylated peptides. Not all peaks are labeled for clarity, but virtually all represent tensin 1 peptides. Each tensin 1 mass analysis generated similar base peak chromatograms with the exception of variability caused by site mutation of tensin 1 peptide **VEAVFSYGPEK**. The site of mutation (F302A) is underlined. Lowercase “c” denotes Cys residues that are carbamidomethylated.

2.5.3.2 Confident Peptide Identification

Data analysis for each tensin 1 condition experiment was guided using an OMSSA search of the MS data against a tensin 1 WT sequence database (gi 187954597, Accession BC140942.1, NCBI). Several factors were used to confidently identify tensin 1 peptides in each data set. First, parts-per-million (ppm) mass accuracy of suspect precursor ions was calculated. Second, correct charge state assignment from monoisotopic resolution in MS1 scans was investigated. Third and most importantly, all MS/MS spectra for unmodified and post-translationally modified peptides were manually validated to ensure confident identification. In addition, relative quantitation of peptides was completed manually in order to determine the degree of PTM. Relative quantitation uses the signal intensity of two standard peptides, angio and vaso, in known concentrations to approximate the abundance of peptides within the same mass spectrometric analysis.

The ability to measure MS1 scans using image currents in the Orbitrap or FT-ICR cell yields routine mass accuracies less than 5 ppm. Because mass accuracy should be relatively consistent in each analysis, ppm mass accuracies can be compared between known standard peptides and those from tensin 1. The mass accuracies of standard and sample peptides are expected to agree within a few ppm. The addition of standard angio and vaso peptides with theoretical monoisotopic masses of $[M+3H]^{+3}$: 432.8998 m/z and $[M+3H]^{+3}$: 475.8846 m/z respectively, allow

for the determination of ppm mass accuracies against which to gauge all others

using **Equation 2.1**.

$$(2.1) \text{ ppm mass error} = \frac{(\text{experimental } m/z) - (\text{theoretical } m/z)}{(\text{theoretical } m/z)} \times 10^6$$

In this fashion, phosphorylated and carbamidomethylated tensin 1 peptide HLGSGSVVPGsPcLDR with an experimental $[M+3H]^+$ at 592.2974 m/z and theoretical $[M+3H]^+$ at 592.2694 m/z yields the following mass defect:

$$\text{ppm mass error} = \frac{(592.2704 \text{ } m/z) - (592.2694 \text{ } m/z)}{(592.2694 \text{ } m/z)} \times 10^6 = + 1.7 \text{ ppm}$$

Generally, ion m/z values that exhibit mass accuracies < 5 ppm are considered high probability identifications on properly calibrated high resolution mass spectrometers.

Detection of distinct isotopic distributions, another advantage afforded by the use of high resolution mass analyzers, is used to disregard peptides which differ from expected isotope patterns. Because the calculated monoisotopic mass for a suspect peptide corresponds to only the ^{12}C isotope of a prescribed charge state, any ions that fall within the mass XIC window that do not match the expected isotopic distribution can be ignored temporarily in the search for the correct ion. However, due to small mass differences that can result from naturally occurring amino acid mutations and certain post-translational modifications, it is important to interrogate ions close in mass in the event that they are structurally related.

The most crucial method for unambiguous identification of peptides using MS is using MSⁿ for peptide structure interrogation. Imperative to the validity of this analysis is the assignment of fragment ions in MS/MS spectra that match the peptide precursor. Predicted fragment ions (b- and y-type for CAD and c- and z•-type for ECD, ETD), in addition to those resulting from neutral losses, form the backbone of spectral assignments. **Figure 2.8** displays singly phosphorylated tensin 1 peptide HLGSGSVVPGsPcLDR (tensin 1 F302A with calyculin A treatment) interrogated using CAD fragmentation. Typically, CAD fragmentation results in optimal dissociation of short peptides ($z=+2$) that are not modified with labile moieties. Labile post-translational modifications like phosphorylation cause localized fragmentation at the site of PTM attachment, resulting in the presence of only a few fragment ions available for confident amino acid sequence and PTM assignment. **Figure 2.8** demonstrates this common occurrence and the lack of complete sequence coverage that ensues. Neutral losses of phosphoric acid, characteristic of collision-based fragmentation, complicate spectra and can make sequence determinations and PTM site mapping more difficult; however, this is not always the case. In the given example, the assignment of a few fragment ions localizes the phosphate group to S1314 based on the knowledge that mammalian phosphorylated residues only include Ser, Thr and Tyr. Complete fragmentation at all residues is not observed. This effect can be even more dramatic and often results in the inability to localize PTMs to specific residues, a central goal of proteomic studies.

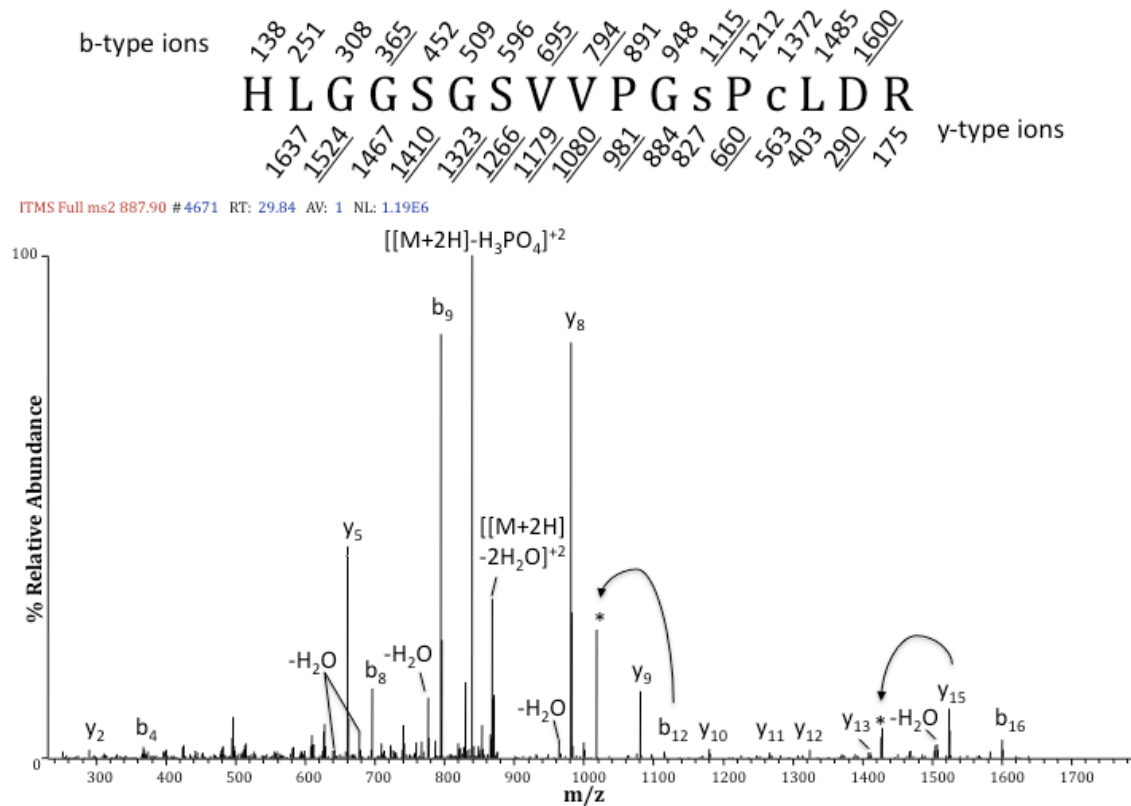


Figure 2.8 CAD MS/MS spectrum of singly phosphorylated and carbamidomethylated tensin 1 peptide, HLGGSVVPGsPcLDR. Dominant loss of phosphoric acid (-98 Da) from the precursor ion is a hallmark of collisional-activation of phosphopeptides and often prevents complete fragment ion sequence coverage. Fragment ions detected are labeled within the spectrum and underlined in the peptide sequence. Ions labeled with * represent those resulting from neutral loss of phosphoric acid. Neutral water losses are labeled. For reference, retention time (RT), number of spectra averaged (AV) and normalized ion counts (NL) are provided.

Interrogation of the same singly phosphorylated, carbamidomethylated tensin 1 peptide HLGGSVVPGsPcLDR (tensin 1 F302A with calyculin A treatment) using ETD fragmentation results in nearly complete fragment ion coverage (**Figure 2.9**). Radical-driven fragmentation does not promote loss of labile modifications to allow successful PTM site identification. The stipulation that ETD requires peptides $z \geq +3$ and m/z values generally between 300 and 800 m/z leaves out a portion of peptides produced by tryptic digestion. (25) Therefore, both CAD and ETD were

required to provide complimentary data sets for optimum sequence and PTM site identification.

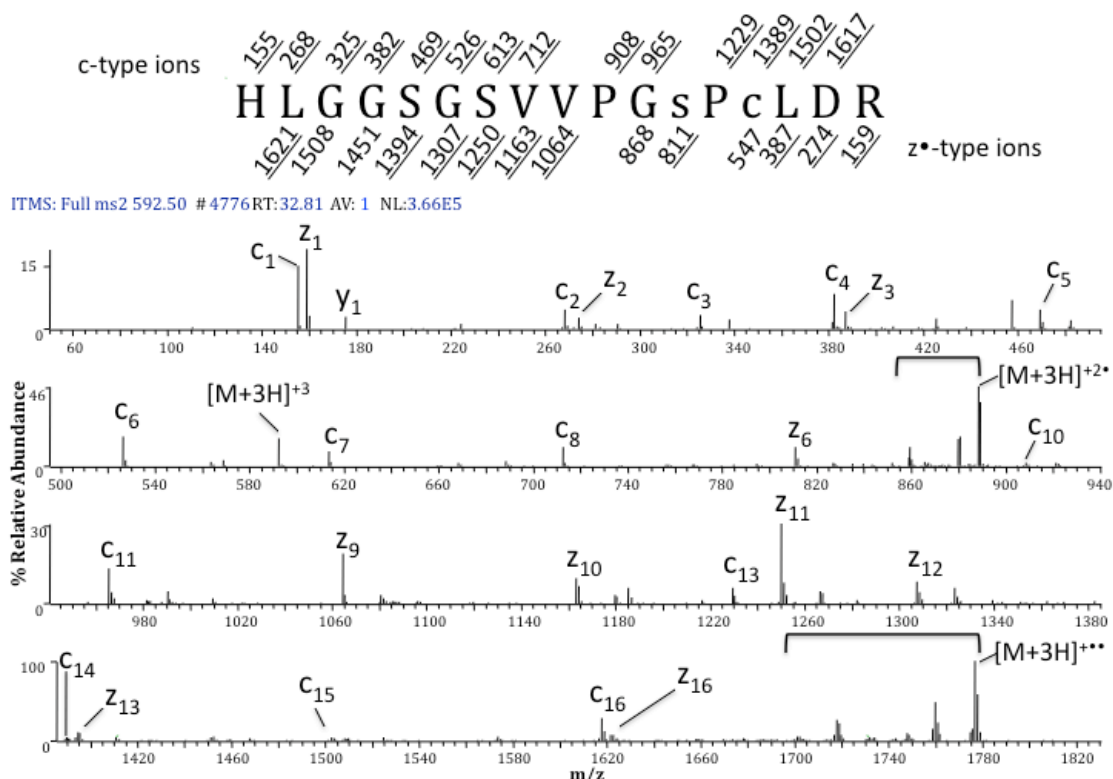


Figure 2.9 ETD MS/MS (75 ms reaction time) of singly phosphorylated and carbamidomethylated **HLGGSVVPGsPcLDR**. Fragment ions detected are labeled within the spectrum and underlined in the peptide sequence. Note that no loss of phosphoric acid is observed. Nearly all fragment ions expected are detected and demonstrate modification at S1314. Fragment ions proximal to Pro residues are not expected because backbone cleavage is not achieved upon electron transfer. Bracketed ions correspond to neutral losses from labeled reduced charge species.

2.5.3.3 Relative Peptide Quantitation

The ion signal for each standard peptide was determined using an extracted ion chromatogram (XIC) centered ± 0.005 m/z about the calculated ^{12}C monoisotopic mass. These XICs were used to identify all charge states present. The integration of ion intensities equaling the summed peak area yields the approximate ion signal for the concentration of peptide added. **Figure 2.10** displays XICs for

angio and vaso used to determine the approximate ion signal for 250 fmol of

peptide. The ion intensity of 250 fmol was calculated by taking the average of angio

and vaso intensities as in **Equation 2.2**.

$$(2.2) \text{ Average 250 fmol area} = \frac{\sum \text{Area}_{250 \text{ fmol angio}} + \sum \text{Area}_{250 \text{ fmol vaso}}}{2}$$

Therefore:

$$\text{Average 250 fmol area} = \frac{2.7E9 + 1.6E9}{2} = 2.15E9$$

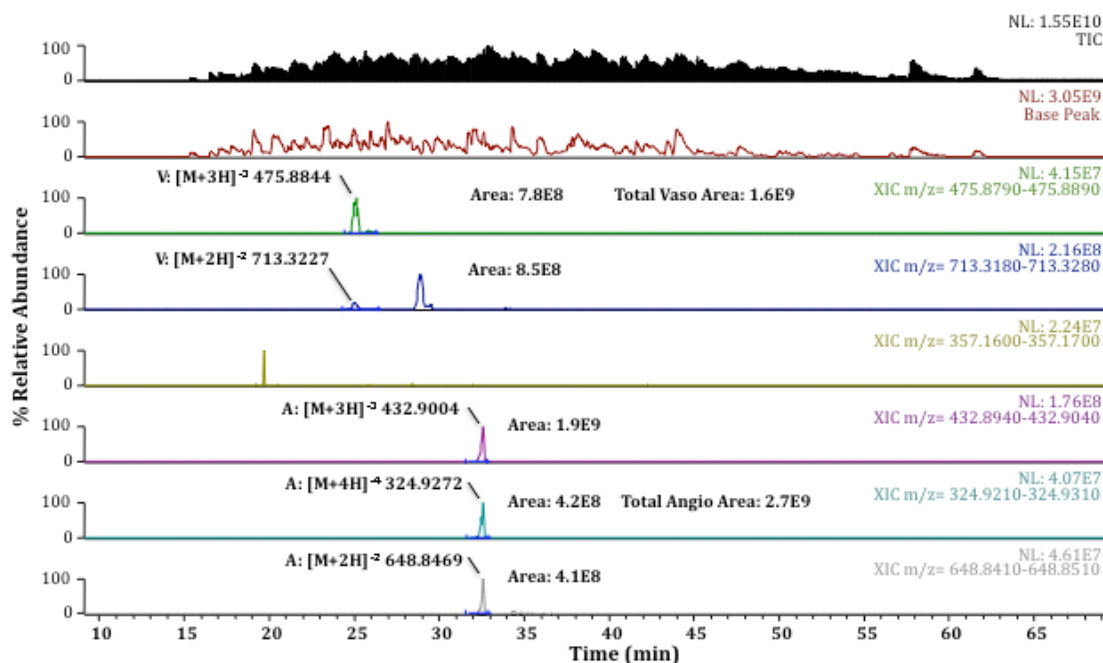


Figure 2.10 XICs of standard peptides angiotensin (A) and vasoactive intestinal peptide (V) used for relative quantitation.

Tensin 1 relative quantitation was determined using the average 250 fmol signal intensity (**Equation 2.2**) and ion intensities for individual peptides calculated using ± 0.005 m/z window XICs. These concentrations were calculated using

Equation 2.3 and the summed peak areas for peptides including each charge state and all molecular forms present, including \pm PTMs (Ser, Thr, Tyr), carbamidomethylation (Cys) and oxidation (Met).

$$(2.3) \frac{250 \text{ fmol}}{2.15E9} = \frac{x \text{ fmol}}{\sum Area_{tensin1 \text{ peptide}}}$$

Again using tensin 1 peptide HLGGSVVPGSPcLDR (tensin 1 F302A with calyculin A treatment) as an example for illustrating peptide identification and quantitation, XICs were constructed for unmodified and singly phosphorylated species (**Figure 2.11A**). High resolution MS1 scans acquired in the Orbitrap provided monoisotopic resolution (**Figure 2.11B,C**) and <5 ppm mass accuracy. Using **Equation 2.3**, estimation of the molar amount of HLGGSVVPGSPcLDR present in the sample is possible.

$$\frac{250 \text{ fmol}}{2.15E9} = \frac{x}{(1.45E10 + 8.72E9)}$$

$$x = 2.7 \text{ pmol}$$

By extending this calculation to all tensin 1 peptides identified, it was estimated that on average, 5% analyses of tensin 1 corresponded to approximately 2-5 pmol. Therefore, nearly consistent concentrations of purified protein for each tensin 1 condition were equal to at 50-100 pmol. This large amount of purified protein demonstrates the successful purification using the S-tag system, a method we instituted for all subsequent tensin 1 analyses.

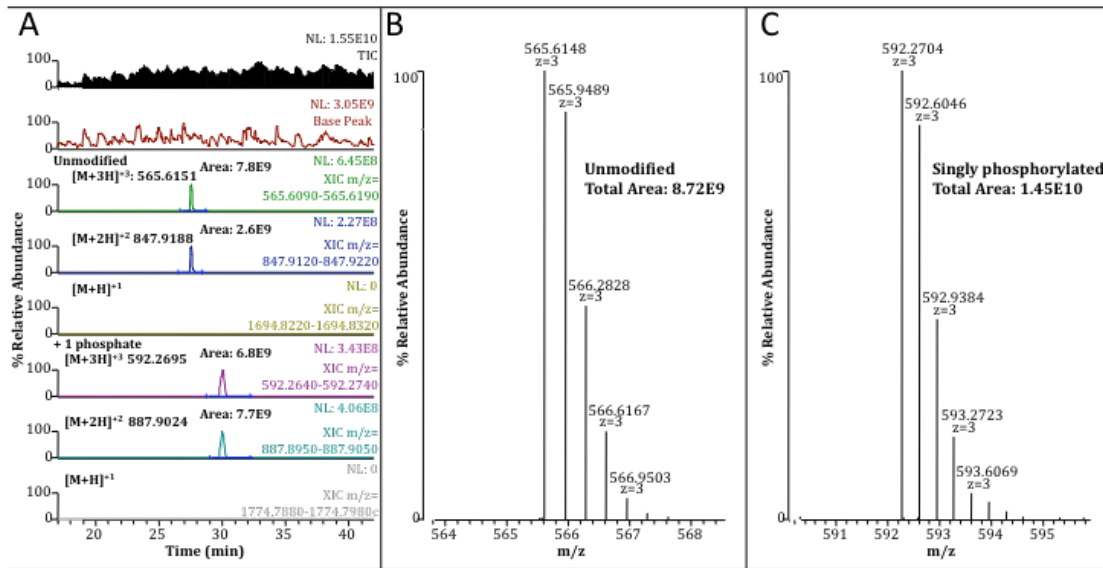


Figure 2.11 XICs and MS1 data for carbamidomethylated tensin 1 peptide HLGGSGSVVPGSPcLDR. (A) XICs and MS1 data for (B) unmodified and (C) singly phosphorylated forms are presented. Peak areas derived from XICs allow relative quantitation, whereas MS1 data displays the expected charge state and isotopic distribution.

Equation 2.4 makes it possible to relatively quantify the amount of PTM for any given peptide.

$$(2.4) \% PTM = \frac{\sum Area_{All\ PTM\ -containing\ species}}{\sum Area_{All\ unmodified\ forms} + \sum Area_{All\ PTM\ -containing\ species}} \times 100\%$$

While whole protein analysis would afford the ability to see which PTMs are present simultaneously, current methods are not available to analyze proteins as large as tensin 1. Therefore, the determination of % PTM yields the most valuable insight regarding PTMs at any single modified residue. Using **Equation 2.4** it is determined that S1314 is phosphorylated at approximately 62.4%.

$$(2.4) \% PTM = \frac{1.45E10}{8.72E9 + 1.45E10} \times 100\% = 62.4\% \text{ phosphorylated}$$

Compilation of all data pertaining to tensin 1 WT and F302A yielded remarkable sequence coverage. The single mutation F302A had no effect on sequence coverage. Due to high sample uniformity between all four tensin 1 conditions, sequence coverage was essentially the same for all analyses. As seen in **Figure 2.12**, nearly 94% of all residues (1629 of 1735 total) were identified. Of 362 total phosphorylatable residues (223 Ser, 91 Thr, 48 Tyr), only 9 were not identified using reverse phase HPLC and mass spectrometric analysis as described. Detection of 98% of all phosphorylatable residues for a protein as large as tensin 1 is significant and highly uncommon, especially when using only a single proteolytic digestion.

S-tagged-TNS1: KETAAAKFERQHMS...LEF

```

1  MSVSRTMEDS CELDLVYVTE RIIAVSFPST ANEENFRSNL REVAOMLKSK HGGNYLLFNL
61 SERRPDITKL HAKVLEFGWP DLHTPALEKI CSICKAMD'TW LNADPHNVV LHNKGNRGRI
121 GVVIAAYMHY SNISASADQA LDRFAMKRFY EDKIVPIGQP SORRYVHYFS GLLSGSIKMN
181 NKPLFLHHVI MHGIPNFESK GGCRPFRLRY OAMQPVYTS IYNIPGDSQT SVCITIEPGL
241 LLKGDILLKC YHKKFRSPAR DVIFRVQFHT CAIHDLG'VF GKEDLDDAFK DDRFPEYK'V
301 EFVFSYGPEK IQGMEHLENG PSVSDYNTS DPLIRWDSYD NFSGHRDDGM EEVVGHTOGP
361 LDGSLYAKVK KKDSLHGSTG AVNATRPTLS ATPNHVEHTL SVSSDSGNST ASTKTDK'TE
421 PVPGASSATA ALSPQEKREL DRLLSGFGL REKOGAMYHT OHLRSRPAGG SAVPSSGRHV
481 VPAQVHVNGG ALASERETDI LDD'ELPNDG HSAGSMG'TLS SLDGV'TNTSE GGYPEALSPL
541 TNGLDKSYPM EPMVNGGGYP YESASRAGPA HAGHTAPMRP SYSAQEGLAG YOREGPHPAW
601 POPVT'TSHYA HDP'SGMFRSO SFSEAE'POLP PAPV'RGSSR EAVQ'RLNSW OOO'OOO'OOO
661 RPPPRQ'QERA HLES'LVASRP SPOPLAETPI PSLPEFPRAA SQOEIEOSIE TLNMLMLDLE
721 PASAAAPLHK SQSV'GAWPG ASPLSSOPLS GSSR'OSHPLT QSRSGYIPSG HSLGTPEPAP
781 RASLESVPPG RSYS'PYD'YOP CLAGPNODFH SKSPASSSLP AFLPTTHSPP GPOPPASLP
841 GLTAOPLLS P KEATSDPSRT PEEEPLNLEG LV'AH'RVAGVQ AREKQPAEPP APLRRRAASD
901 GQYENOSPEA TSPRS'PGVRS PVQC'VSP'ELA LTIALNPGGR PKEPHLHSYK EAFEEMEGTS
961 PSSPPP'SGVR SPPGLAKTPL SALGLKPHNP ADILLHPTGV TRRRIOPEED EGKVVVRLSE
1021 EPRSYVESVA RTAVAGPRAO DSEPKSFSAP ATOAYGHEIP LRNGTLGGSF VSPSPLSTSS
1081 PILSADSTSV GSF'PSGESSD OGPRTPTOPL LESGFRSGSL GQPSPSAORN YQSSSPLPTV
1141 GSSYSSPDYS LQHFSSSPES QARAQFSVAG VHTVPGSPOA RHRTVGTNTP PSPGFGWRAI
1201 NPSMAAPSSP SLSHHOMMGP PGTGFHGSTV SSPOSSAATT PGSPSLCRHP AGVYQVSGHL
1261 NKVATTPGSP SLGRHPGAHQ GNLASGLHSN AIASPGSPSL GRHLGGSGSV VPGSPCLDRH
1321 VAYGGYSTPE DRRPTLSROS SASGYOAPST PSFPVSPAYY PGLSSPATSP SPDSAAFROG
1381 SPTPALPEKR RMSVGD'RAGS LPNYATINGK VSSPVASGMS SPSGGSTVSF SHTLPDFSKY
1441 SMPDN'NPETR AKVKFVODTS KYWKPEISR EQAIALLKDO EPGAFIIRDS HSFRGAYGLA
1501 MKVSSPPPTI MOONKKGDMT HELV'RHF'IE TGPRGVKLG CPNEPNFGSL SALVYQHSII
1561 PLALPCKLVI PNRDPTDESK DSSGPANSTA DLLKQGAACN VLFVNSVDME SLTGPOAISK
1621 ATSETLAADP TPAATIVHFK VSAOGITLTD NORKLFFRRH YPLNTVTFCD LDPQERKWMK
1681 TEGGAPAKLF GFVARKOGST TDNACHLFAE LDPNQASAI VNFVSKVMLN AGQKR

```

Figure 2.12 Sequence coverage map of tensin 1. Underlined residues correspond to those included in detected peptides. The sequence is numbered with disregard to the tag sequence and begins with the initiating Met. Sequence coverage for each condition analyzed was virtually identical.

Extensive manual validation of all MS/MS spectra using visual inspection for unmodified and post-translationally modified peptides resulted in the identification of many phosphorylation sites in untreated tensin 1 F302A. Surprisingly, comparison of data from untreated tensin 1 WT resulted in 50 identical Ser and Thr phosphorylation sites (**Figure 2.13**). These sites were identified using collective CAD and ETD MS/MS data sets as illustrated by the example peptide in **Figures 2.8 & 2.9**.

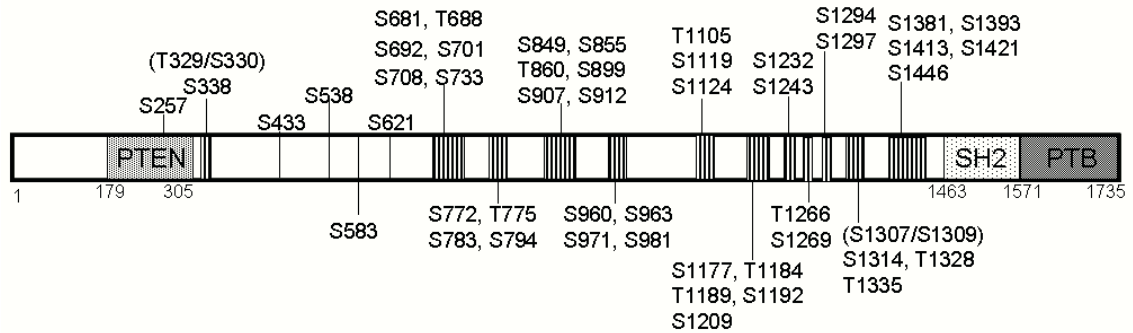


Figure 2.13 Phosphorylated residue distribution on tensin 1 WT and F302A without calyculin A treatment. Sites in parentheses indicate peptides that, upon CAD and ETD fragmentation, were not unambiguously assigned to a single residue. Phosphorylation sites S1209, S1232 and S1243 were identified using an chymotrypsin subdigestion.

Similar analysis was conducted for the tensin 1 WT and F302A samples treated with the PP1 calyculin A Ser/Thr phosphatase inhibitor. Resulting from these experiments was the identification of 12 rescued Ser and Thr phosphorylation sites (**Table 2.2**). Again, mass spectral data suggested that the same phosphorylation sites were present in both tensin 1 WT and F302A when treated similarly. Phosphopeptide enrichment using IMAC resulted in the detection of an additional phosphorylated peptide DSSGPANsTADLLK (pS1588) while all other sites were detected in both enriched and unenriched mass analyses.

Peptide Sequence	Phosphorylated Residue
LEFM _s VsR	S2 & S4
DDGMEEVVGHTQGPLDGsLYAK	S364
LLsGFGLER	S445
sRPAGGsAVPSSGR	S465 & S471
GLNsWQQQQQQQQPRPPR	S649
QsHPLTQsR	S756 & S762
LSEEP _R sYVESVAR	S1024
AGsLPNYATINGK	S1400
DSSGPANsTADLLK	S1588 (IMAC exclusive)

Table 2.2 Tensin 1 WT and F302A phosphorylated residues rescued using calyculin A treatment.

Identical phosphorylation sites in tensin 1 WT and F302A were unexpected due to the disruption of PP1 α association upon site mutation. Even though the same

sites were identified overall, it was necessary to address the evidence suggested by the electrophoretic shift changes under SDS-PAGE separation of the two tensin 1 constructs (**Figure 2.4**). It was initially thought that different overall phosphorylation sites on the tensin 1 constructs caused reduced electrophoretic migration. However, extensive phosphorylation site analysis suggests that this is incorrect. An alternate explanation of the SDS-PAGE results stems from varying amounts of phosphorylation (% PTM) per site between conditions as opposed to completely unique PTM sites. A comparison of % PTM values between tensin 1 WT and F302A samples addressed this question. For simplicity, an initial % PTM comparison was made using only the phosphorylation sites on tensin 1 prior to residue 740. This was directed by evidence from an experiment identical to that represented by **Figure 2.4** but used truncated tensin 1 WT and F302A, residues 1-740. **Table 2.3** suggests that there is not an outstanding difference in % PTM for any single site located within in this region, though peptides including single sites S364, S538 and S621 and all three phosphorylated S681, T688, and S692 sites represent those with approximately 2-fold changes in phosphorylation. These changes, which would certainly affect electrophoretic mobility, are suspected to represent limited phosphatase activity as a result of PP1 α binding inhibition with tensin 1 F302A.

Peptide Sequence	Modified Site	WT	F302A
LEFM _s V _s R	S2 & S4	1.8%	1.5%
IQGMEHLENGPSVSDYN(TS)DPLIR	T329/T330	1.4%	1.4%
WD _s YDNFSGHR	S338	100%	100%

DDGMEEVVGHTQGPLDGsLYAK	S364	4.5%	<1%
TDKTDEPVPGASSATAALsPQEK	S433	14.1%	13.4%
LLsGFGLER	S445	15.2%	9.9%
sRPAGGsAVPSSGR	S465 & S471	1.0%	3.3%
ETDILDDELPNQDGHSAGSMGTLSSLD GVTNTSEGGYPEALsPLTNGLDK	S538	11.2%	25.1%
SQsFSEAEPQLPPAPVR	S621	66.7%	38.5%
GLNsWQQQQQQQQPRPPPR	S649	1.6%	2.0%
AHLESLVASRPsPQPLAETPIPsLPEFPR	S681, T688 & S692	8.9%	3.7%
AAAsQQEIEQsIETLNMLMLDLEPASAAAPLHK	S701 & S708	7.4%	7.8%
SQsVPAGAPGSPLSSQPLSGSSR	S733	9.3%	9.4%

Table 2.3 Relative abundances of phosphorylation in tensin 1 WT and F302A (1-740) with calyculin A treatment.

2.5.4 Differential Isotopic Labeling Phosphorylation Enrichment Analysis

2.5.4.1 Advantages of Differential Labeling and IMAC

Due to the observed changes in phosphorylation in the 1-740 truncation comparison, a differentially labeled phosphorylation analysis was implemented in order to determine changes over the entire tensin 1 sequence under the same conditions. This experiment allows simultaneous analysis of two identical proteolytic digests (tensin 1 WT and F302A) by incorporating identifiable mass tags. Isotopic labeling is preferred because even though differentially labeled peptides with identical amino acid sequences co-elute during an HPLC gradient, separate identification is achieved by added mass differences. Therefore, MS1 scans simultaneously report both ions with respective added masses and abundances required for relative abundance determination.

Coupling differential labeling with phosphopeptide enrichment yields several advantages. In general, the most obvious benefit afforded by phosphopeptide-specific enrichment is the ability to increase the experimental dynamic range by removing unmodified peptides. Unmodified counterpart peptides are often highly abundant and therefore are preferentially chosen for data-dependent MS/MS analysis. By implementing IMAC, the efficiency of phosphorylation-centered analysis is heavily increased. IMAC requires separate derivatization of side chain (Asp, Glu) and C-terminal carboxylic acids due to similar electrostatic interactions with the Fe(III)-imidodiacetate resin. By incorporating isotopic labeling within the necessary esterification reaction, two potential issues are resolved at once. Differential labeling of tensin 1 WT and F302A was achieved using D₀ methanolic-HCl (+14 Da/acid) and D₃ methanolic HCl (+17 Da/acid), respectively. Prior to IMAC enrichment, aliquots of each tensin 1 WT and F302A labeled samples were separately mass analyzed to assure complete derivatization of acidic residues. Peptide concentrations were determined for each to allow mixing of both labeled digests in equal proportions before enrichment.

In order to monitor activity specific to PP1 α , the differential labeling enrichment was completed on the tensin 1 constructs untreated with calyculin A. Otherwise, PP1 α activity would be obstructed. It is indeed possible that other Ser/Thr phosphatase-mediated changes in phosphorylation are represented in the resulting data. However, due to the specificity of calyculin A for PP1 and the known

disruption of PP1 α binding to tensin 1 WT, we hypothesize that the results should correlate strongly to the action of this single phosphatase.

2.5.4.2 Data Analysis

Relative quantitation of all peptides was done similarly to methods previously described. Efficiency of the enrichment was gauged by the recovery of two phosphorylated standard peptides, angio II phosphate (DRVpYIHPF) and STsLVEGR. Both were added in 250 fmol concentrations. Comparison of peak areas in a 250 fmol analysis completed prior to enrichment with the isotopic labeling enrichment resulted in recoveries of approximately 30% and 100% for DRVYIHPF and STsLVEGR, respectively. Average peptide recoveries for tensin 1 phosphopeptides are likely to fall between these values.

Modified tensin 1 peptides were manually identified using XICs centered ± 0.005 Da about the ^{12}C isotopic mass (**Figure 2.14A**). Isotopic distributions for heavy (D_3 , F302A) and light (D_0 , WT) peptides were validated based on expected charge states and monoisotopic masses of phosphorylated peptides identified in previous analyses (**Figure 2.14B**). In addition, OMSSA data searches were applied in order to identify any potential new phosphorylation sites, though none were identified. In order to easily recognize differences in phosphorylation between tensin 1 F302A and WT, " D_3/D_0 " ratios were calculated for each phosphopeptide identified as demonstrated in **Equation 2.5**. For simplicity, the D_3/D_0 ratio was determined using the ^{12}C isotopic peak area for only the most abundant charge state

in each case. As is evident in **Figure 2.14**, however, the ratio is stable regardless of the examined charge state.

$$(2.5) \frac{D_3}{D_0} = \frac{^{12}\text{C isotopic peak area of tensin F302A phosphorylated peptide}}{^{12}\text{C isotopic peak area of tensin WT phosphorylated peptide}}$$

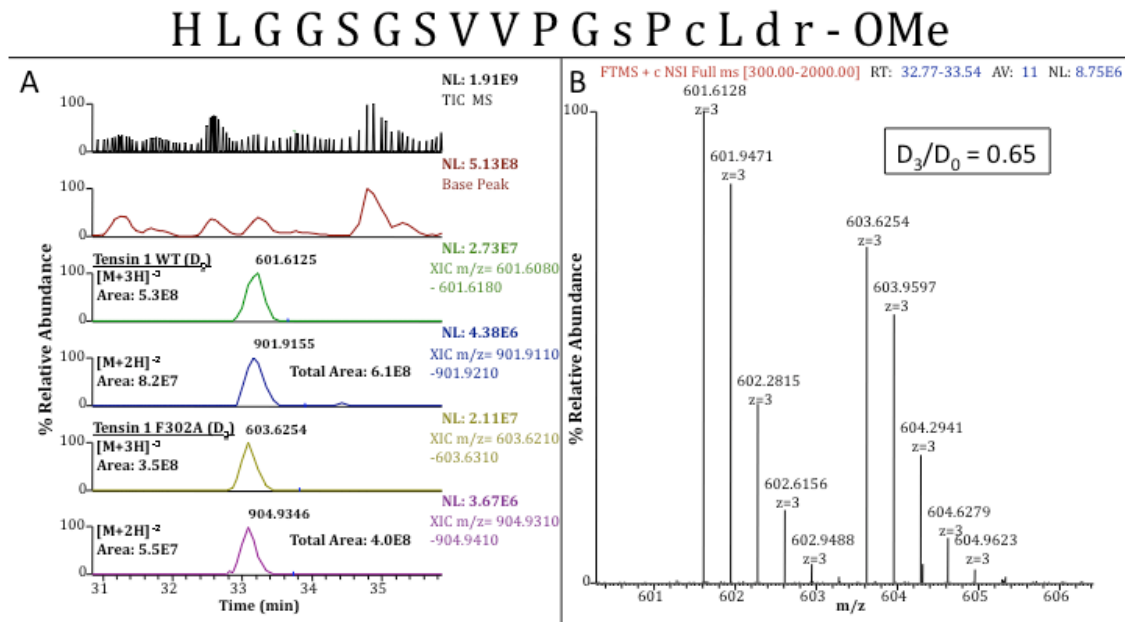


Figure 2.14 D_3/D_0 ratio determination for HLGGSVVPGsPcLDR. (A) XICs for tensin 1 WT and F302A ^{12}C monoisotopic peaks. The D_3/D_0 ratio is determined from the areas under the ^{12}C monoisotopic peak using the most abundant charge state for each tensin 1 F302 and WT. Though $z=+3$ was used for this peptide, $+2$ values demonstrate that ratios remain constant over other charge states. (B) By averaging over the MS1 scans during the complete elution time for both coeluting peptides, abundance comparisons are illustrated.

While it might be suggested from the initial SDS-PAGE data (**Figure 2.4**) that large differences in phosphorylation exist, surprisingly, the large majority of D_3/D_0 ratios were close to 1. This demonstrates that there was essentially no change in phosphorylation because the peptides were initially mixed at a ratio of 1. **Table 2.4** highlights the 7 most interesting phosphorylated peptides; that is, those that

displayed the largest D_3/D_0 ratios. One would expect that disruption of the interaction between tensin 1 and PP1 α resulting from a F302A site mutation would result in higher levels of phosphorylation at sites specific to PP1 α . It stands to reason then, that phosphorylated peptides with the highest D_3/D_0 ratios (>2) are most likely PP1 α targets. Interestingly, peptides that fall within this category are interspersed throughout the entire tensin 1 sequence. The highest D_3/D_0 ratio corresponds to the peptide containing pS338, a phosphorylated residue very close to the site of interaction with PP1 α at F302. We hypothesize that this residue is a very likely substrate for PP1 α , among others that appear to be affected to a lesser degree. Likely due to the size of tensin 1, the full length crystal structure has not been determined. This information would provide spatial proximity information to determine whether these sites are not only close to each other, but also to the region where PP1 α binds to tensin 1.

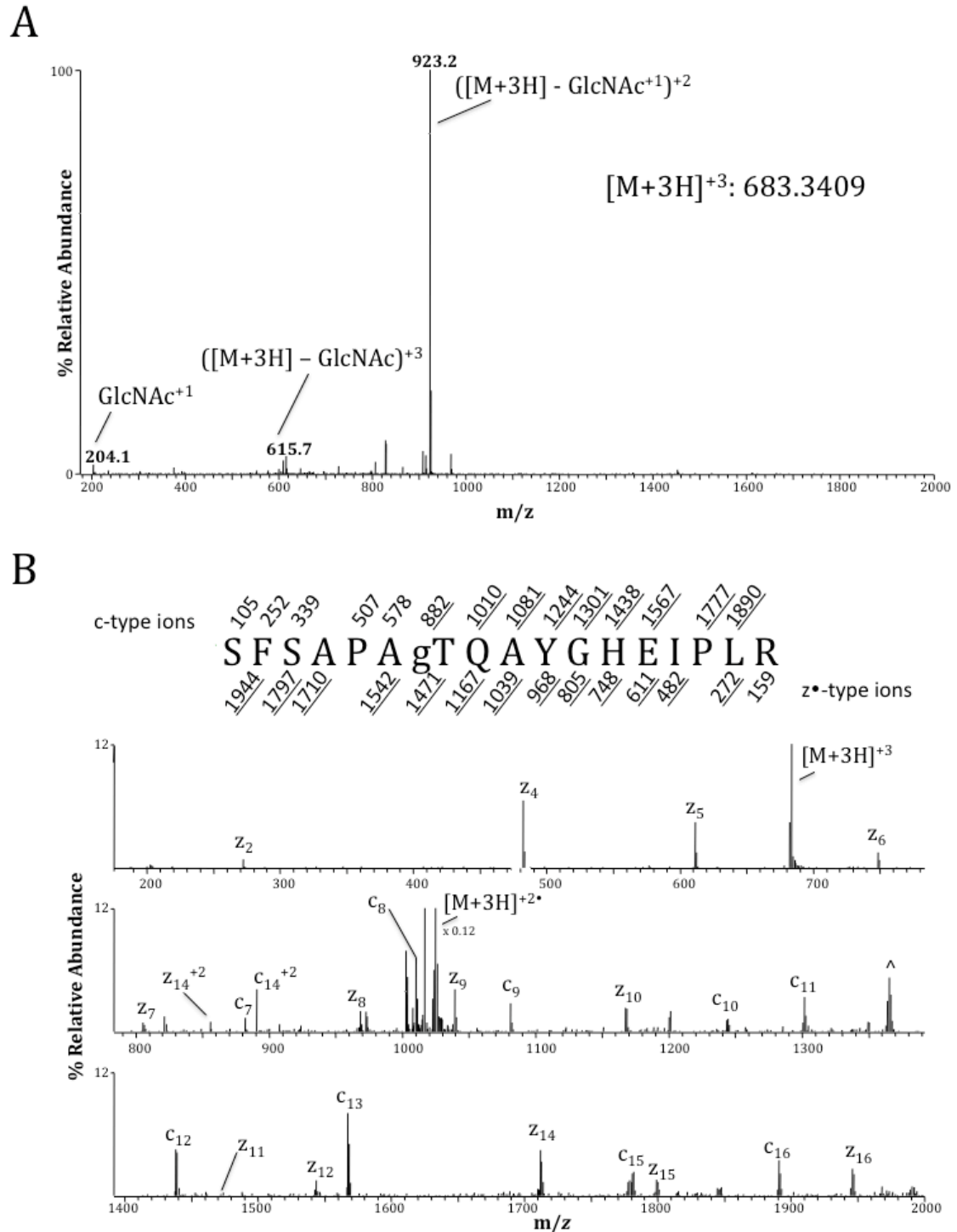
Residues	Peptide Sequence	D_3/D_0
336-346	WDSYDNFSGHR	13
1200-1248	AINPSMAAPSSPSLSHHQMMGPPGTGFHGSTVSSPQSSAATTPGSPSLCR	5.3
813-851	SPASSSLPAFLPTTHSPPGPQQPPASLPGLTAQPLLSPK	3
860-875	TPEEEPLNLEGLVAHR	2.5
1392-1397	MSVGDR	2.3
1379-1389	QGSPTPALPEK	2.1
255-260	FRSPAR	2.1

Table 2.4 Phosphorylated tensin 1 peptides with the greatest D_3/D_0 ratios.

4.5.4.3 *O*-GlcNAcylation of Tensin 1 WT and F302A

Upon further inspection of mass spectral data, it was determined that both tensin 1 WT and F302A were modified with a different PTM, *O*-linked β -N-acetylglucosamine (*O*-GlcNAc). This enzymatically-regulated monosaccharide

located on Ser and Thr residues was first discovered by Gerald Hart in 1984 and is notoriously difficult to localize using conventional collision-based mass spectrometry. More extensive background information on the signaling aspects and difficulties in characterization using MS can be found in Chapter 3: Post-translational Modification Analysis of Homeodomain-Interacting Protein Kinase 1, a collaborative effort with the Hart laboratory. Briefly, gas phase lability of the modification upon collision in addition to ion suppression of modified peptides by highly abundant unmodified peptides makes characterization of this modification nearly impossible; that is, until the recent application of ETD MS. While the ion suppression issue has yet to be routinely resolved, ETD has circumvented the issue of *O*-GlcNAc's lability under collisional activation. Fragmentation using CAD results in three characteristic ions: 1) a singly charged GlcNAc oxonium ion at m/z 204.1, 2) the corresponding reduced charge precursor ion following oxonium ion loss, $[(M+3H)-GlcNAc^{+1}]^{+2}$ and 3) an ion resulting from neutral loss of the PTM, $[(M+3H)-GlcNAc]^{+3}$ (**Figure 2.15A & 2.16A**). Due to the lability of *O*-GlcNAc, very few fragment ions exist to facilitate PTM site assignment. Often, any fragment ions that do exist exhibit loss of *O*-GlcNAc, which always prevents site assignment. However, interrogation with ETD retains the modification and produces fragment ions that allow unambiguous site identification (**Figure 2.15B & 2.16B**).



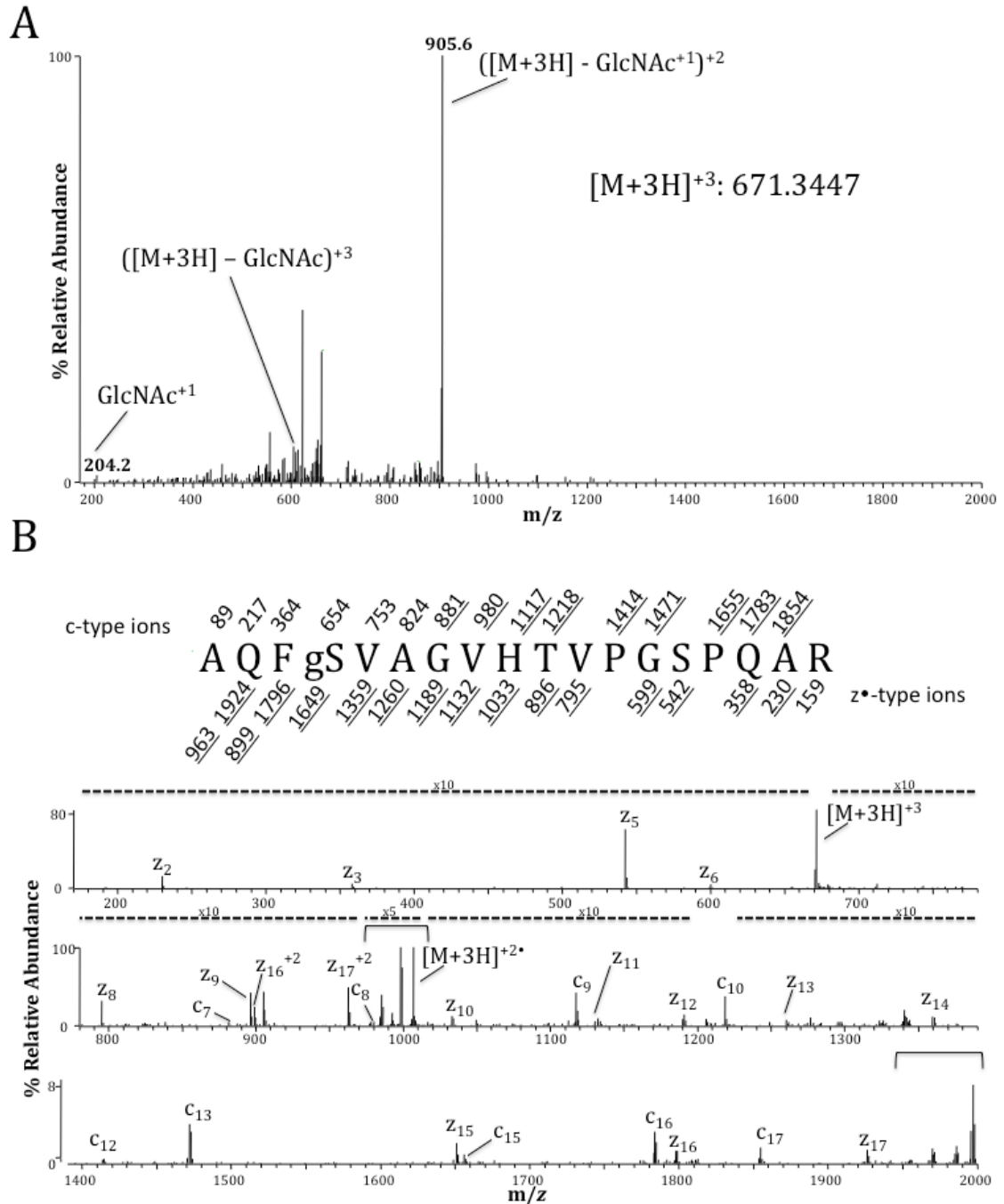


Figure 2.16 (A) CAD and (B) ETD MS/MS spectra of *O*-GlcNAcylated tensin 1 peptide, AQFgSVAGVHTVPGSPQAR. Detected fragment ions are labeled within the spectrum and underlined in the peptide sequence. Site assignment is made possible only using ETD MS.

Currently, there is no reproducible and efficient method for the enrichment of *O*-GlcNAc-modified peptides. This is due to the lack of functional groups that are

amenable to enrichment. Therefore, identification of these post-translationally modified peptides is still dictated by current chromatographical and mass spectrometric methods. Despite the inability to enrich for this PTM, we were able to characterize differences in presence of the modification (**Table 2.5**). *O*-GlcNAcylated residue T1052 was confidently identified in tensin 1 WT and F302A with calyculin A treatment. However, *O*-GlcNAcylated S1167 was unambiguously identified only in tensin 1 F302A with calyculin A.

Peptide Sequence	<i>O</i> -GlcNAc Site	Tensin 1 WT	Tensin 1 F302A
SFSAPAgTQAYGHEIPLR	T1052	x	x
AQFgSVAGVHTVPGSPQAR	S1167		x

Table 2.5 Tensin 1 *O*-GlcNAcylation results with calyculin A treatment.

2.5.5 Tyrosine Phosphorylation Enrichment Analysis of tensin 1 WT

Prior to the current phosphorylation analysis, tensin 1 was already known to be a phosphoprotein. A large number of the reported phosphorylation sites were located on Tyr residues. However, we identified no pTyr residues in the initial tensin 1:PP1 α investigation. We initially thought this was attributed to characteristically low abundance of Tyr phosphorylation as compared to Ser and Thr phosphorylation. Because of the 70 phosphorylated Ser and Thr residues present on tensin 1 in this study, it was generally unsurprising that no Tyr phosphorylation was identified, even with IMAC. If low amounts of pTyr did exist, highly abundant pSer and pThr-containing peptides would dominate data dependent MS/MS analysis. However, it was determined that no detectable levels of

pTyr were present after immunoblotting against pTyr using the 4G10 antibody

(Figure 2.17A).

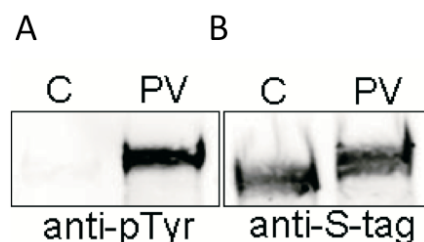


Figure 2.17 Immunoblotting tensin 1 against phosphorylated pTyr using (A) control (C) and (B) peroxyvanadate (PV)-treated tensin 1 WT.

In order to complete a comprehensive analysis of tensin 1 phosphorylation in HEK293 cells, we purified S-tagged tensin 1 WT in the presence of Tyr phosphatase inhibitor peroxyvanadate to rescue all existing Tyr phosphorylation prior to lysis. Immunoblotting against pTyr with peroxyvanadate shows largely increased amounts of tensin 1 Tyr phosphorylation. Following an initial data-dependent analysis of the peroxyvanadate treated sample, IMAC enrichment and interrogation using CAD allowed the identification of 10 pTyr sites (**Table 2.6**). Efficiency of this enrichment was gauged using two phosphorylated standard peptides. Calculated recoveries of DRVpYIHPF and STsLVEGR were 80% and 2%, respectively. Recovery of STsLVEGR proved difficult to quantify due to the coeluting, esterified tensin 1 peptide (RMsVGDR) that falls within the same prescribed XIC window (m/z 464.7104 and 464.7072, respectively). Because all spectra and XICs contain both species, a definitive recovery value is difficult to attain. Recovery of DRVYIHPF is likely a more accurate value to estimate the other phosphopeptide recoveries.

Peptide Sequence	Phosphorylation Site	Novel PTM?
DDGMEEVVGHTQGPLDGSLyAK	Y366	Yes
QGAMyHTQHLLR	Y458	No
EGPHPAWPQPVTTSHyAHDPSGMFR	Y609	No
SGyIPSGHSLGTPEPAPR	Y766	No
SyVESVAR	Y1025	No
SFSAPATQyGHEIPLR	Y1055	No
HPAGVyQVSGLHMK	Y1254	Yes
HVAyGGYSTPEDR	Y1323	Yes
HVAYGGySTPEDR	Y1326	Yes
ySMPDNSPETR	Y1440	No

Table 2.6 Phosphotyrosine results following treatment with peroxyvanadate and enrichment using IMAC.

At the time of discovery, four of the identified pTyr sites were previously unreported. **Figure 2.18** displays the distribution of phosphorylated Tyr residues in full length tensin 1. Isotopically labeled analyses were not carried out on Tyr phosphorylation because of the assumed irrelevance in regards to PP1 α phosphatase activity.

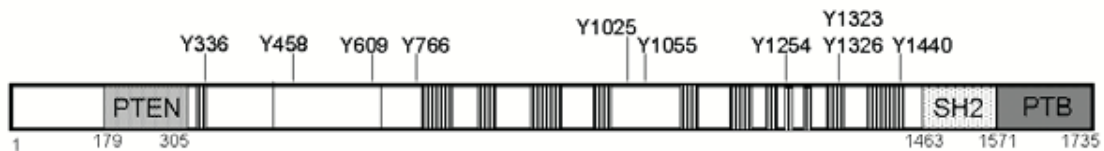


Figure 2.18 Distribution of Tyrosine Phosphorylation on peroxyvanadate-treated tensin 1 WT.

Overall, tensin 1 was determined to be a highly modified phosphoprotein on Ser, Thr and Tyr residues also modified with *O*-GlcNAc on Ser and Thr. A complete summary of the PTMs identified on tensin 1 is found in **Table 2.7**. These results were obtained by complementary CAD and ETD data. In addition, phosphopeptide enrichment using IMAC afforded not only an IMAC-exclusive site at S1588, but also the ability to unambiguously identify the site of several PTMs. This was made

possible because the coeluting, unrelated, and isobaric species that complicated many MS/MS spectra were removed.

Residues	Peptide Sequence	Site(s)	Observed m/z	z	Observed Mass (Da)	Calculated Mass (Da)	IMAC	Site Assignment
1-5	LEFM _s V _s R	S2 & S4	524.7313	+2	1048.4553	1048.4533	x	CAD
255-260	FR _s PAR	S257	414.2000	+2	827.3927	827.3924	x	CAD
311-335	IQGMEHLENGPSVSDYN(ts)D PLIR	T329/S330	951.1046	+3	2851.2992	2851.2968		CAD
336-346	WD _s YDNFSGHR	S338	488.5179	+3	1463.5391	1463.5375		CAD/ETD
347-368	DDGMEEVVGHTQG _s PLDG _s LYAK	S364	828.0518	+3	2482.1408	2482.1211	x	CAD/ETD
347-368	DDGMEEVVGHTQG _s PLDG _s LYAK	Y366	828.0494	+3	2482.1336	2482.1210	x	CAD
415-437	TDKTDEVPVGASSATAAL _s PQEK	T433	794.0356	+3	2380.0922	2380.0915		CAD/ETD
443-451	LL _s GFGLER	S445	536.2664	+2	1071.5255	1071.5234		CAD
454-464	QGAM _s YHTQH _s LR	Y458	560.2568	+3	1678.7558	1678.7519	x	CAD
465-478	sRPAGG _s AVPSSGR	S465 & S471	683.3170	+2	1365.6267	1365.6270		CAD
520-546	SSLDGVTNTSEGGYPEAL sPLTNGLDK	S538	934.7656	+3	2802.2822	2802.2717		CAD
567-593	AGPAHAGHTAPMRPSY _s A QEGLAGYQR	S583	723.0908	+4	2889.3413	2889.3251	x	ETD
594-618	EGPHPAWPQPVTTSY _s HAH DPSGMFR	Y609	975.4429	+3	2924.3141	2924.2975	x	CAD
619-635	SQ _s FSEAE _s QLPPAPVR	S621	640.6354	+3	1919.8916	1919.8898		CAD/ETD
646-665	GLN _s WQQQQQQQQPRPPR	S649	837.0676	+3	2509.1882	2509.1844		CAD
670-698	AHLES _s LVASRP _s PQPLA _s EtPIP sLPEFPR	S681, T688 & S692	1144.8926	+3	3432.6632	3432.6408	x	CAD
699-730	AA _s QQEIEQ _s IE _s TLNMLMLDLEP ASAAAPLHK	S701 & S708	1203.8938	+3	3609.6668	3609.6617		CAD/ETD
731-754	SQ _s VPGAWPGASPLSSQP LSGSSR	S733	807.3809	+3	2420.1281	2420.1241		CAD
755-763	Q _s HPLTQ _s R	S756 & S762	378.5083	+3	1133.5103	1133.5099		CAD
764-781	SGYIPSGHSLGTPEPAPR	Y766	634.9660	+3	1902.8834	1902.8745		CAD
764-781	SGYIPSGH _s LgtPEPAPR	S772 & T775	670.9663	+3	2010.8843	2010.8723	x	ETD
782-791	A _s LESVPPGR	S783	546.7588	+2	1092.5103	1092.5085		CAD
792-812	SY _s PYD _s YQCLAGPNQDFHSK	S794	866.0383	+3	2596.1003	2596.0847	x	CAD
813-851	SPASSSLPAFLPTTHSPGPGQQP PASLPLGTAQPLL _s PK	S849	988.7647	+4	3952.0369	3952.0255		CAD
852-875	EAT _s DPSRtPEEPLNLEGLVAHR	S855 & T860	727.0952	+4	2905.3589	2905.3384	x	ETD
897-914	AA _s DGQYENQ _s PEATSPR	S899	663.2723	+3	1987.8023	1987.8029		CAD
897-914	AASDGQYENQ _s PEATSPR	S907	663.2720	+3	1987.8014	1987.8029		CAD
897-914	AASDGQYENQ _s PEATSPR	S912	663.2720	+3	1987.8014	1987.8029		CAD
951-970	EAFEEMEGT _s PS _s PPPSGVR	S960 & S963	1125.9376	+2	2250.8679	2250.8662		CAD
971-977	sPPGLAK	S971	749.3597	+1	749.3597	749.3593		CAD
978-1002	TPL _s ALGLKPHNPADILLHPTGVTR	S981	675.3657	+4	2698.4409	2698.4440		ETD
1018-1031	LSEEP _s YVESVAR	S1024	567.9329	+3	1701.7841	1701.7844		ETD

1024-1031	SyVESVAR	Y1025	509.7353	+2	1018.4633	1018.4605	x	CAD
1046-1062	SFSAPAgTQAYGHEIPLR	O-GlcNAc @ T1052 (in WT & F302A)	683.3418	+3	2048.0108	2048.0083		ETD
1046-1062	SFSAPATQAYGHEIPLR	Y1055	651.6498	+3	1952.9348	1952.9267	x	CAD
1105-1116	tPTQPLLESGFR	T1105	485.2441	+3	1453.7177	1453.7088	x	CAD/ETD
1117-1129	SGsLQQP _s PSAQR	S1119 & S1124	723.2989	+2	1445.5905	1445.5822	x	CAD
1164-1181	AQFsVAGVHTVPGPQAR	OGlcNAc @ S1167 (in F302A)	671.3448	+3	2012.0198	2012.0195		ETD
1164-1181	AQFSVAGVHTVPGsPQAR	S1177	630.3079	+3	1888.9091	1888.9065		CAD
1184-1198	tVGTNtPPsPGFGR	T1184, T1189 & S1192	781.3320	+2	1561.6567	1561.6448	x	CAD
1199-1225	AINPSMAAPsPSLSHH	S1209	595.2682	+3	1783.7900	1783.7833		ETD
1226-1248	HGSTVs _s PQSSAATTPGSPSLCR	S1232	784.6840	+3	2352.0374	2352.0281		ETD
1226-1248	HGSTVSSPQSSAATTPG _s PSLCR	S1243	784.6833	+3	2352.0353	2352.0281		CAD/ETD
1249-1262	HPAGV _y QVSLH _{NK}	Y1254	793.8819	+2	1586.7565	1586.7475		CAD
1263-1274	VATtPG _s PSLGR	T1266 & S1269	658.7900	+2	1316.5727	1316.5647	x	CAD
1275-1302	HPGAHQGNLASGLHSNAIA _s PG _s PSLGR	S1294 & S1297	716.8337	+4	2864.3129	2864.2990	x	ETD
1303-1319	HLGG(sGs)VVPGsPCLDR	S1314 & (S1307/S1309)*	618.9246	+3	1854.7592	1854.7600		CAD
1320-1332	HVAyGGYSTPEDR	Y1323	766.3189	+2	1531.6305	1531.6213		CAD
1320-1332	HVAYGGySTPEDR	Y1326	766.3184	+2	1531.6295	1531.6213		CAD
1320-1332	HVAYGGYstPEDR	T1328	511.2117	+3	1531.6205	1531.6213		CAD/ETD
1333-1338	RptLSR	T1335	412.2155	+2	823.4237	823.4186	x	CAD
1379-1389	QGsPTPALPEK	S1381	602.7842	+2	1204.5611	1204.5609		CAD
1392-1397	RMsVGDR	S1393	450.6911	+2	900.3749	900.3757		CAD
1398-1410	AGsLPNYATINGK	S1400	693.3274	+2	1385.6475	1385.6461		CAD
1411-1439	VSSPVASGMSSPSGGSTVSFSHTLPDFSK	S1413	974.1098	+3	2920.3148	2920.3070		CAD
1411-1439	VSSPVASGMs _s PSGGSTVSFSHTLPDFSK	S1421	974.1082	+3	2920.3100	2920.3070		CAD
1440-1450	ySMPDNsPETR	Y1440	688.7671	+2	1376.5269	1376.5188		CAD
1440-1450	YSMPDNsPETR	S1446	688.7635	+2	1376.5197	1376.5188		CAD
1581-1594	DSSGPANsTADLLK	S1588	749.3502	+2	1497.6931	1497.6833	x	CAD

Table 2.7 Complete PTM identifications resulting from this comprehensive analysis. Modified peptide residue positions, sequences, and PTM sites are listed. Mass accuracy is represented by the theoretical and calculated precursor masses as determined from the most abundant charge state detected. PTM sites unambiguously identified using IMAC data are designated by "x." The type of fragmentation used for PTM site assignment is listed as well. Peptides Asp1199-His1225 and His1226-Arg1248 were generated from a chymotrypsin digestion.

2.5.6 *In vitro* Kinase Assays

The exceptionally large number of tensin 1 phosphorylation sites exist with varied neighboring amino acid sequences. This suggests a highly complex mechanism that likely includes many different kinases and phosphatases. Though we believe PP1 α is part of this system, it is unlikely that it exclusively dephosphorylates tensin 1 because many sites are still present after inhibition of PP1 α association with tensin 1. In order to gain some understanding of where many of these phosphorylation sites originated, we focused on the majority of the Ser and Thr phosphorylation sites identified that exist in Ser/Thr-Pro motifs. This suggests PTM by Pro-directed kinases, a group of enzymes that includes glycogen synthase kinase 3 (GSK3), mitogen-activated protein kinases (MAPK) and cyclin dependent kinases (CDKs). While it was not clear initially which group(s) of kinases were responsible, it was observed that at least one kinase associated with tensin 1 strongly enough to survive S-tag purification. Kinase activity was evident when purifying S-tagged tensin 1 WT in Golden buffer but not when using RIPA buffer even though both methods purified nearly equal amount of tensin 1. This suggested that the associated kinase does not maintain interaction with tensin 1 when subjected to more stringent detergent conditions. *In vitro* incubation of tensin 1 and associated proteins with γ [³²P] ATP resulted in the incorporation of additional phosphate in as little as 30 min (**Figure 2.19**).

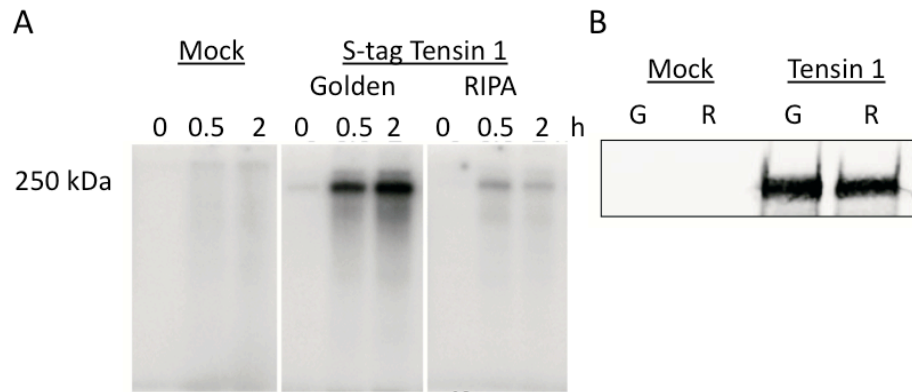


Figure 2.19 (A) Phosphor-image (^{32}P) analysis shows active phosphorylation in the presence of ATP during incubation after lysis with (B) equal amounts of tensin 1 purified.

Efforts to identify the active kinase required specific kinase inhibitors to those types of Pro-direction kinases mentioned previously. An *in vitro* assay measured reductions in phosphorylation. The treatments and inhibitors used were: 1) mock buffer (control), 2) GSK3 β inhibitor II, 3) SB220025 (MAPK inhibitor), and 4) staurosporine (broad inhibitor). Dramatic changes in the decrease of tensin 1 phosphorylation resulted only in the presence of SB220025 and staurosporine (**Figure 2.20**). Nearly equal decreases in ^{32}P signal intensity when using these inhibitors suggests that p38 MAPK is a major kinase that phosphorylates tensin 1.

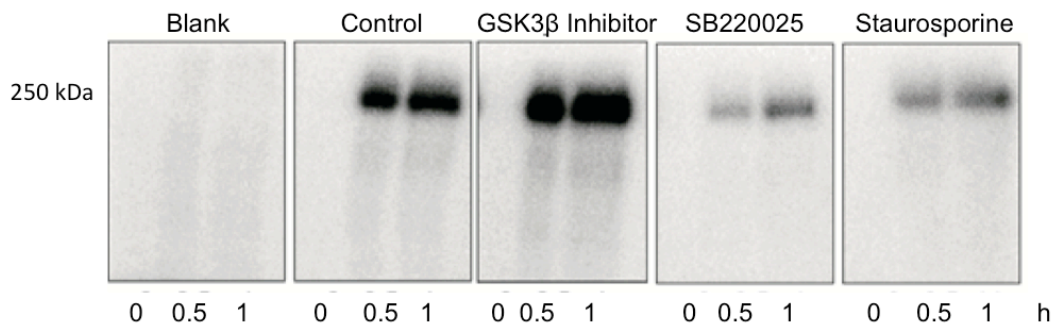


Figure 2.20 *In vitro* effect of kinase inhibitors on tensin 1 phosphorylation.

Quantitation of the ^{32}P signal intensity changes are presented in **Figure**

2.21A. Immunoblotting using anti-p38 MAPK ensures that endogenous p38 MAPK is indeed co-purified with S-tagged tensin 1 (**Figure 2.21B**). We show *in vitro* that p38 MAPK is able to phosphorylate tensin 1 extensively. The hypothesis that a single kinase p38 MAPK is responsible for such a large portion of total phosphorylation is surprising due primarily to the sheer number of modified residues on the protein.

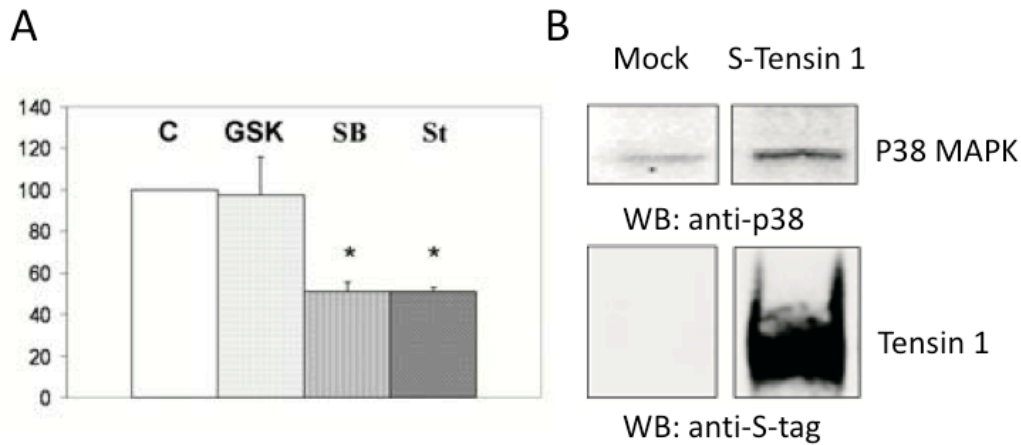


Figure 2.21 (A) Quantitation of phosphorylation reduction in tensin 1 and (B) Co-IP of p38 MAPK using anti-p38 MAPK and anti-S-tag to recognize tensin 1. Inhibitors are as follows: control (C), GSK3 β inhibitor (GSK), SB220025 (SB) and staurosporine (St).

2.5.7 DLC-1 Binding Assays

The numerous phosphorylation sites on tensin 1 located away from the N- and C-termini may suggest one or a combination of two situations. First, tensin 1 could play a role in an extremely complex signaling network involving many phosphorylation sites that each trigger unique downstream signaling effects without involving tensin 1 termini. Second, the extreme phosphorylation could

provide a possible electrostatically favored binding platform through which proteins may interact. Because tensin 1 is already known to interact with the tumor suppressor protein DLC-1, we investigated how p38 MAPK-mediated phosphorylation affects DLC-1 binding.

Purified recombinant p38 MAPK was used to *in vitro* phosphorylate S-tagged tensin 1 purified from HEK293 cells and resulted in 2.5-fold higher phosphorylation (**Figure 2.22A**). Expected background levels of phosphorylation were present in the control. While incubation with lambda phosphatase was initially implemented to provide a larger contrast in phosphorylation levels, kinase incubation post-lambda phosphatase treatment resulted in no ³²P incorporation. This could be due to lack of successful phosphatase removal despite washes with RIPA buffer or the inability of p38 MAPK to phosphorylate completely unmodified tensin 1. Regardless, the 2.5-fold difference in phosphorylation between control and p38 MAPK-treated tensin 1 resulted in a significant 80% decrease in DLC-1 binding (**Figure 2.22B,C**). These results suggest that *in vitro*, interaction between tensin 1 and DLC-1 is phosphorylation-dependent and is regulated in part by p38 MAPK activity.

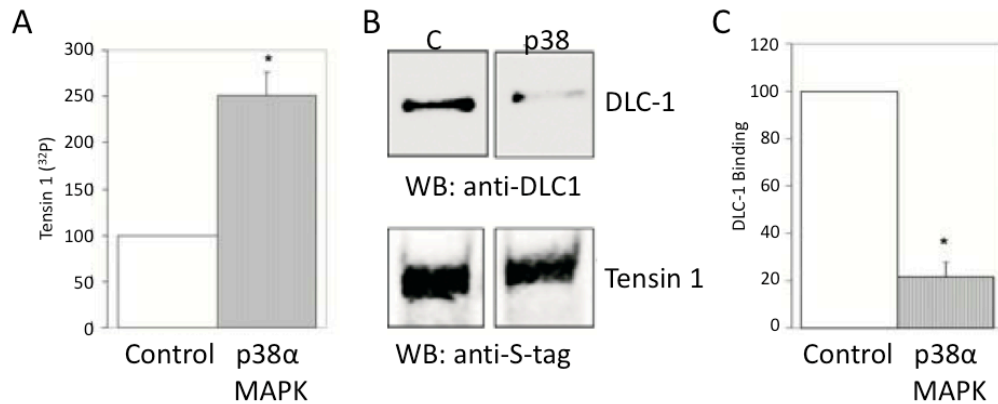


Figure 2.22 The relationship between DLC-1 binding and tensin 1 phosphorylation. (A) Quantitation of tensin 1 phosphorylation levels without (control) and with recombinant p38 MAPK. (B) Western blotting against anti-DLC1 demonstrates reduced DLC-1 binding with p38 MAPK phosphorylation. (C) Quantitation of the amount of DLC-1 associating with p38 MAPK-phosphorylated tensin 1.

2.5.8 Hyperosmotic Sorbitol Treatment

Tensin 1 phosphorylation by p38 MAPK was next manipulated *in vivo*.

HEK293 cells expressing S-tagged tensin 1 were treated with the sugar alcohol sorbitol to induce hyperosmotic stress and p38 MAPK activation. p38 MAPK activation is characterized by phosphorylation of two Thr and Tyr residues within a Thr-Xxx-Tyr sequence (26). Based on previous experiments, we expected that sorbitol treatment should result in higher phosphorylation and decreased association with DLC-1 for S-tag-purified tensin 1.

Indeed, sorbitol treatment of HEK293 cells did result in p38 MAPK phosphorylation and activation (**Figure 2.23A**). As expected, Ser and Thr phosphorylation of tensin 1 results from *in vivo* p38 MAPK activation and decreased amounts of DLC-1 binding to S-tagged tensin 1 isolated by S-protein agarose (**Figure 2.23B-D**). Increases in phosphorylation are evident in both band intensity

and the characteristic decrease in electrophoretic mobility. It is important to note that DLC-1 used for these binding experiments was expressed in untreated HEK293 cells and used for incubation prior to Western blot analysis. This source of DLC-1 guarantees that changes in binding between the two proteins are caused by changes in tensin 1 phosphorylation exclusively. Otherwise, it is possible that altered MAPK activity resulting from sorbitol treatment could also post-translationally modify DLC-1 and cause altered binding tendencies. These results coincide with previous *in vitro* data that characterize DLC-1 interaction disruption caused by p38 MAPK phosphorylation of tensin 1.

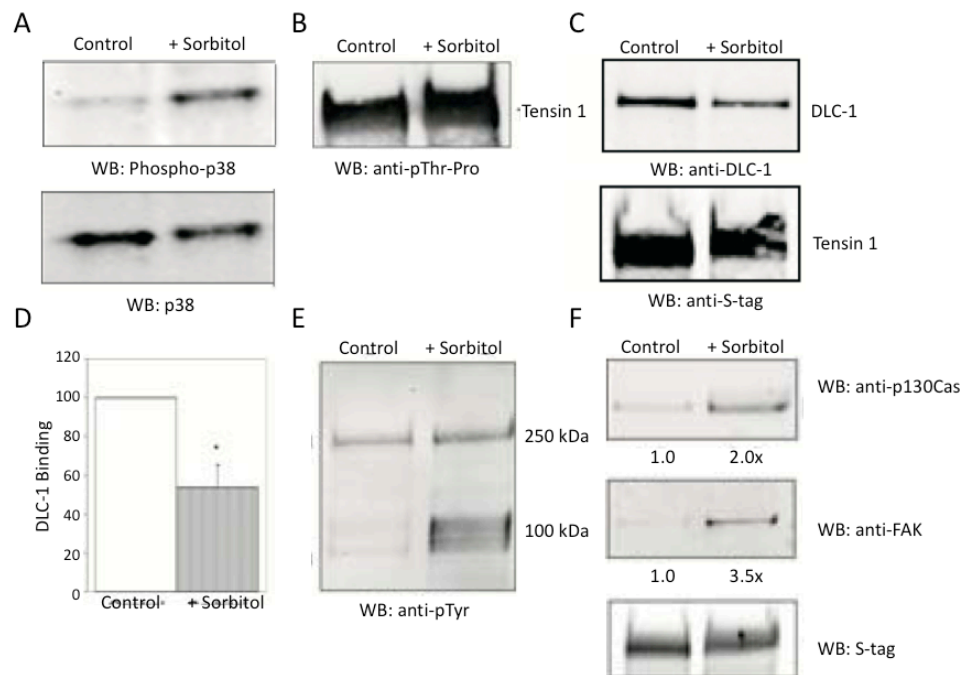


Figure 2.23 Sorbitol treatment and the effect on tensin 1 interacting proteins. (A) Addition of sorbitol results in increased phosphorylated (activated) p38 MAPK, (B) higher amounts of pThr-Pro in tensin 1, (C) higher amounts of p38 MAPK-mediated phosphorylation and reduced DLC-1 association with tensin 1. (D) Quantitation of DLC-1 binding displays a reduction upon sorbitol treatment. (E) Sorbitol treatment results in varied levels of pTyr phosphorylation in co-purified focal adhesion proteins, measured using 4G10. (F) In particular, sorbitol treatment increases tensin 1 association with p130Cas and FAK.

Lastly, because focal adhesion-based protein interactions are often regulated by Tyr phosphorylation, Tyr phosphorylation was investigated under sorbitol treatment as well. Using anti-pTyr antibody 4G10 to blot against endogenous proteins that naturally interact with S-tagged tensin 1, it was evident that levels of Tyr phosphorylation increased in several unidentified proteins (~100 kDa) while nearly identical amounts of pTyr remained on tensin 1 (~250 kDa) (**Figure 2.23E**). To interrogate alternate protein binding under sorbitol treatment for proteins known to be involved in focal adhesions with tensin 1, p130Cas and focal adhesion kinase (FAK), immunoblot analyses were used (**Figure 2.23F**). Though similar band intensities when blotted against the S-tag (tensin 1) suggest equal amounts of tensin 1 purified under both conditions, the abundance of interacting proteins p130Cas and FAK significantly increases, suggesting further that changes in tensin 1 phosphorylation mediate protein interactions.

2.6 Conclusions

These results present the first successful comprehensive analysis of phosphorylation sites in tensin 1. The ability to characterize phosphorylation and *O*-GlcNAcylation sites on this very large protein demonstrates the high efficiency of PTM analysis using proteolytic digestion, high resolution and tandem mass spectrometry online with HPLC separations. These methods, combined with phosphopeptide enrichment by IMAC that incorporated isotopic labeling strategies, allowed for the identification of almost all 1,735 amino acids constituting the tensin

1 primary sequence. Identified PTMs include 50 pSer/pThr phosphorylation sites without calyculin A treatment, an additional 12 pSer/pThr sites with calyculin A treatment and 10 pTyr sites with peroxyvanadate treatment. Two novel *O*-GlcNAcylation sites add to an incredible total of 74 PTMs on tensin 1. These results indicate that at least 19% of all Ser/Thr and 21% of all Tyr residues on tensin 1 are substrates for phosphorylation. As of the date of publication in *Molecular & Cellular Proteomics* (Aug 2010), 41 pSer/pThr, 2 pTyr and both *O*-GlcNAcylation sites were previously unreported.

Several factors contributed to the success of this PTM analysis. S-tag purification afforded virtually exclusive isolation of tensin 1 and increased the dynamic range of the experiment accordingly. Many low abundance PTM identification experiments are limited due to the presence of numerous contaminating proteins that compete for data dependent MSⁿ analysis. Consequently, we were able to identify phosphorylation sites at levels below 1% relative abundance. The sensitivity and confidence in PTM site assignment afforded by high resolution MS1 and tandem mass spectrometric analyses incorporating both CAD and ETD fragmentation allowed unambiguous site mapping of all but two sites. Complementary ETD and CAD data validated by visual inspection and manual quantitation granted superior confidence in PTM characterization. To lack the ability for confident characterization aside from statistical values often puts high throughput phosphoproteomic analyses in question. Phosphopeptide enrichment resulted in the exclusive identification of pS1588 and made unambiguous site

identification possible for several sites by decreasing the complexity of the analysis to primarily phosphopeptides.

Aside from the analytical achievements of this study, changes in phosphorylation on tensin 1 WT and F302A suggest that PP1 α dissociation from tensin 1 affects the amount of phosphorylation on seven sites to a significant degree (**Table 2.4**). These results were determined through the use of PP1 phosphatase inhibitor, calyculin A. An interesting aspect of these results is the location of the phosphorylation sites. With few exceptions, all modified side chains exist near the central region of tensin 1 outside of known binding domains. This suggests that although sequence homology may not attribute distinct domains to this region, it is indeed part of a regulatory pathway and is highly post-translationally modified.

In vitro and *in vivo* experiments have provided insight into specific p38 MAPK activity on tensin 1. Phosphorylation by p38 MAPK has direct effects in regards to DLC-1 binding and association with other focal adhesion proteins FAK and p130Cas. All three proteins were previously known to interact in tensin 1-containing focal adhesions (21). Specifically, we show that DLC-1 binding is reduced upon sorbitol-induced phosphorylation by p38 MAPK. Because we know interaction of DLC-1 with tensin 1 can stably exist, we believe it is at least in part regulated by p38 MAPK-mediated phosphorylation at Ser/Thr-Pro sequences but also by selective PP1 α -mediated dephosphorylation. This hypothesis agrees with previous data that demonstrates DLC-1 binding is favored with tensin 1:PP1 α

interaction. It would stand to reason then, that DLC-1 interacts optimally only when PP1 α dephosphorylates sites catalyzed by p38 MAPK. While the unique functions of all identified phosphorylation sites are not known, one function is likely to include mediation of interactions with other focal adhesion proteins. The observation that both p130Cas and FAK more favorably interact with tensin 1 with increased p38 MAPK activity supports this model.

Site mutation at R1488 has been shown to disrupt DLC-1 binding to tensin 1. The closest Ser/Thr-Pro phosphorylation site identified is located at S1446. This is relatively far based on primary sequence alone, but the three-dimensional structure could allow this or virtually any of the Ser/Thr-Pro sites to be close enough to electrostatically disrupt DLC-1 interactions. This suggests that indeed, extensive phosphorylation in the central region of tensin 1 plays a role in stability and regulation of protein:protein interactions. We hypothesize that changes in electrostatic interactions that result from such extensive phosphorylation are the basis for these effects. Unfortunately, because tensin 1 is extremely large, the crystal structure has not yet been determined. Elucidation of this structure could help to understand how specific sites interact with regions on the surface of tensin 1 to induce changes in protein:protein interactions.

The function of tensin 1 WT and F302A *O*-GlcNAcylation site T1052 and F302A-exclusive site S1167 are still not understood. *O*-GlcNAc is known to function in phosphorylation-based signaling cascades through modification of

phosphoproteins (7). The addition of *O*-GlcNAc near to or on sites known to be phosphorylated can prevent the addition of phosphate during signaling events (7). As a result, *O*-GlcNAc is involved in a complex interplay with phosphorylation. We suspect that *O*-GlcNAcylation of tensin 1 is acting in a similar fashion though the exact mechanism of interplay at work is unknown. Interestingly, both *O*-GlcNAcylation sites also have phosphorylation sites fewer than 10 residues away. This distance falls within the region Burkinaree et al. suggest can still influence phosphate-based signaling mechanisms (8). Hart and coworkers previously determined that *O*-GlcNAc transferase (OGT), the enzyme that catalyzes *O*-GlcNAc attachment to proteins, associates directly with PP1 (27). We suspect that interaction between PP1 α and OGT facilitates *O*-GlcNAc modification of tensin 1, though this mechanism is not understood.

In conclusion, we present the identification of many novel PTMs on tensin 1, a key cell migration regulatory protein that is modified in part by p38 MAPK. Disruption of protein:protein interactions with several other important signaling and focal adhesion-related proteins as a result of changes in phosphorylation demonstrate the importance of these modifications. Analytical mass spectrometry single-handedly allowed the identification of numerous phosphorylation sites. These results contribute to a newly developed hypothesis that tensin 1 acts in conjunction with several other key proteins to serve as a regulatory molecule for focal adhesion stability to control cell migration.

2.7 References

1. Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks, Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P. & Mann, M., *Cell* **2006** (127): 635-648.
 2. Signaling--2000 and Beyond, Hunter, T., *Cell* **2000** (100): 113-127.
 3. Decoding signalling networks by mass spectrometry-based proteomics, Choudhary, C. & Mann, M., *Nat Rev Mol Cell Biol* **2010** (11): 427-439.
 4. Farley, A. R. & Link, A. J. (2009) in *Methods in Enzymology*, ed. Deutscher, R. R. B. a. M. P. (Academic Press, pp. 725-763.
 5. The role of protein phosphorylation in human health and disease., Cohen, P., *European Journal of Biochemistry* **2001** (268): 5001-5010.
 6. Post-translational modifications in signal integration, Deribe, Y. L., Pawson, T. & Dikic, I., *Nature Structural & Molecular Biology* **2010** (17): 666-672.
 7. Cross-talk between GlcNAcylation and phosphorylation: Site-specific phosphorylation dynamics in response to globally elevated O-GlcNAc, Wang, Z., Gucek, M. & Hart, G. W., *Proceedings of the National Academy of Sciences* **2008** (105): 13793-13798.
 8. O-linked [beta]-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress, Butkinaree, C., Park, K. & Hart, G. W., *Biochimica et Biophysica Acta (BBA) - General Subjects. Special Issue: Nucleocytoplasmic Glycosylation* **2010** (1800): 96-106.
 9. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy, Zarling, A. L., Polefrone, J. M., Evans, A. M., Mikesh, L. M., Shabanowitz, J., Lewis, S. T., Engelhard, V. H. & Hunt, D. F., *Proceedings of the National Academy of Sciences* **2006** (103): 14889-14894.
 10. Focal adhesions: What's new inside, Lo, S. H., *Developmental Biology* **2006** (294): 280-291.
 11. Regulation of tensin-promoted cell migration by its focal adhesion binding and Src homology domain 2., Chen, H. & Lo, S. H., *Biochemical Journal* **2003** (370): 1039-1045.
 12. Structure of the PTB domain of tensin1 and a model for its recruitment to fibrillar adhesions, McCleverty, C. J., Lin, D. C. & Liddington, R. C., *Protein Science* **2007** (16): 1223-1229.
 13. Srichai, M. B. & Zent, R. (2010) in *Cell-Extracellular Matrix Interactions in Cancer*, eds. Zent, R. & Pozzi, A. (Springer New York, pp. 19-41.
 14. The lamellipodium: where motility begins, Small, J. V., Stradal, T., Vignat, E. & Rottner, K., *Trends in Cell Biology* **2002** (12): 112-120.
 15. PTEN Gene and Integrin Signaling in Cancer, Tamura, M., Gu, J., Tran, H. & Yamada, K. M., *Journal of the National Cancer Institute* **1999** (91): 1820-1828.
-

16. Tensin, Lo, S. H., The International Journal of Biochemistry & Cell Biology **2004** (36): 31-34.
 17. Molecular characterization of human tensin., Chen, H., Ishii, A., Wong, W. K., Chen, L. B. & Lo, S. H., Biochemical Journal **2000** (351): 403-411.
 18. A reciprocal tensin-3-cten switch mediates EGF-driven mammary cell migration, Katz, M., Amit, I., Citri, A., Shay, T., Carvalho, S., Lavi, S., Milanezi, F., Lyass, L., Amariglio, N., Jacob-Hirsch, J., Ben-Chetrit, N., Tarcic, G., Lindzen, M., Avraham, R., Liao, Y.-C., Trusk, P., Lyass, A., Rechavi, G., Spector, N. L., Lo, S. H., Schmitt, F., Bacus, S. S. & Yarden, Y., Nature Cell Biology **2007** (9): 961-969.
 19. Presence of an SH2 domain in the actin-binding protein tensin, Davis, S., Lu, M., Lo, S., Lin, S., Butler, J., Druker, B., Roberts, T., An, Q. & Chen, L., Science **1991** (252): 712-715.
 20. Association of the Tensin N-terminal Protein-tyrosine Phosphatase Domain with the alpha-Isoform of Protein Phosphatase-1 in Focal Adhesions, Eto, M., Kirkbride, J., Elliott, E., Lo, S. H. & Brautigan, D. L., Journal of Biological Chemistry **2007** (282): 17806-17815.
 21. Tensin1 Requires Protein Phosphatase-1a in Addition to RhoGAP DLC-1 to Control Cell Polarization, Migration, and Invasion, Hall, E. H., Daugherty, A. E., Choi, C. K., Horwitz, A. F. & Brautigan, D. L., Journal of Biological Chemistry **2009** (284): 34713-34722.
 22. Comprehensive Analysis of Phosphorylation Sites in Tensin1 Reveals Regulation by p38MAPK, Hall, E. H., Balsbaugh, J. L., Rose, K. L., Shabanowitz, J., Hunt, D. F. & Brautigan, D. L., Molecular & Cellular Proteomics **2010** (9): 2853-2863.
 23. Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry, Udeshi, N. D., Compton, P. D., Shabanowitz, J., Hunt, D. F. & Rose, K. L., Nature Protocols **2008** (3): 1709-1717.
 24. Ribonuclease S-peptide as a carrier in fusion proteins, Kim, J.-S. & Raines, R. T., Protein Science **1993** (2): 348-356.
 25. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry, Good, D. M., Wirtala, M., McAlister, G. C. & Coon, J. J., Molecular & Cellular Proteomics **2007** (6): 1942-1951.
 26. Mechanism of p38 MAP kinase activation in vivo, Brancho, D., Tanaka, N., Jaeschke, A., Ventura, J.-J., Kelkar, N., Tanaka, Y., Kyuuma, M., Takeshita, T., Flavell, R. A. & Davis, R. J., Genes & Development **2003** (17): 1969-1978.
 27. O-GlcNAc Transferase Is in a Functional Complex with Protein Phosphatase 1 Catalytic Subunits, Wells, L., Kreppel, L. K., Comer, F. I., Wadzinski, B. E. & Hart, G. W., Journal of Biological Chemistry **2004** (279): 38466-38470.
-

3 Post-translational Modification Analysis of Homeodomain-Interacting Protein Kinase 1

3.1 Abstract

The homeodomain-interacting protein kinase (HIPK) family consists of four kinases that were discovered by a yeast two-hybrid screen using NK1.2 homeoprotein as bait. The 130 kDa Ser/Thr kinase HIPK1 has been surprisingly understudied, but has been implicated as a mediator for pro-apoptotic signaling within the Fas-mediated pathway. HIPK1 has also been found to interact with *O*-GlcNAc transferase using a separate yeast two-hybrid screen, which suggests it may be modified with *O*-GlcNAc. In general, *O*-GlcNAc site assignments in all proteins are few compared to the number of known phosphorylation sites because of difficulties in detection using conventional mass spectrometric techniques that use CAD for MS/MS analysis. In order to localize HIPK1 *O*-GlcNAc sites, we used multiple stages of proteolysis and mass analyzed the resulting peptides with ETD, a technique that successfully retains *O*-GlcNAc upon MS/MS analysis. Novel charge state modification strategies were applied to manipulate precursor charge states of *O*-GlcNAcylated peptides that originally were not amenable to analysis using ETD. These methods afforded the characterization of 16 novel phosphorylation sites on HIPK1. In addition, 3 novel *O*-GlcNAc sites were localized to important functional regions in the active kinase and C-terminal binding domains. All three *O*-GlcNAc sites are currently being investigated for regulatory functions towards kinase activity and protein interactions that lead to apoptosis.

3.2 Introduction

3.2.1 *O*-GlcNAc and *O*-GlcNAc Transferase

O-linked- β -N-acetylglucosamine (*O*-GlcNAc or GlcNAc) is a monosaccharide protein PTM that modifies Ser and Thr side chains. This PTM was originally discovered by the Hart laboratory in 1984 (1). While polysaccharide attachment to protein side chains has been studied extensively, the discovery of *O*-GlcNAc revealed a new type of protein glycosylation with several unique characteristics. First, *O*-GlcNAc represents a protein modification composed of a single sugar. This directly contrasts with previously known glycosylation patterns, which can exist as N- or O-linked branched polysaccharide chains of variable length (2). Second, *O*-GlcNAcylation has been found to exist on both nuclear and cytosolic proteins (3). Conversely, N- and O-linked branched glycosylation is found on extracellular and secreted proteins, or luminal domains of membrane proteins (2, 4). Third, *O*-GlcNAc has been implicated as an important signal transduction molecule (2).

The ability of a PTM to facilitate signal transduction requires that the modification is transient. Enzymes that catalyze the addition and removal of a specific PTM provide these transient modification states. Kinases and phosphatases are well-characterized enzymes that constitute a phosphate-dependent signaling network of this type. In a similar fashion, *O*-GlcNAc is enzymatically added and removed from proteins by *O*-GlcNAc transferase (OGT) and *O*-GlcNAc aminidase (*O*-GlcNAcase), respectively (**Figure 3.1A**) (5, 6). The donor molecule for the addition

of GlcNAc is uridine diphosphate-GlcNAc (UDP-GlcNAc), which is synthesized via the hexosamine biosynthetic pathway (HBP) (**Figure 3.1B**) (7).

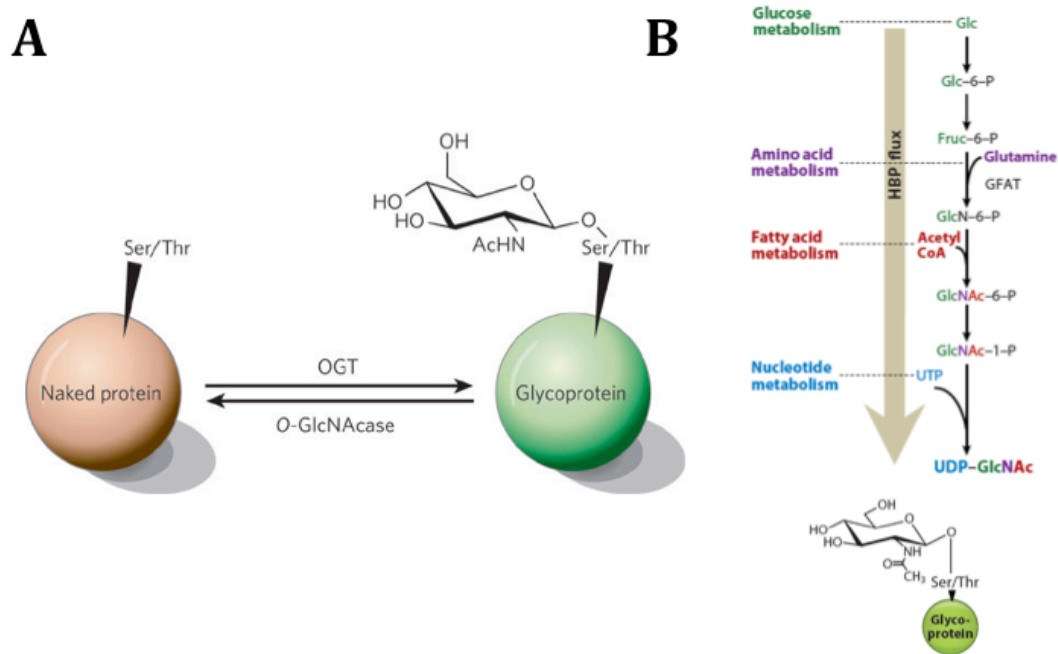


Figure 3.1 Protein GlcNAcylation. (A) Addition and removal of *O*-GlcNAc to Ser or Thr residues is catalyzed by OGT and *O*-GlcNAcase, respectively. (B) UDP-GlcNAc, the donor molecule for GlcNAcylation, is synthesized via the multi-step HBP. Fluxes in HBP can result from increased glucose concentration, which results in increased protein GlcNAcylation. Figures reproduced from (8, 9).

A second requirement for cell signaling molecules is the ability to induce modification by unique stimuli or events. Protein *O*-GlcNAcylation has been shown to be up-regulated by events that include insulin signaling, glucose metabolism and cell cycle progression (2, 9, 10). The importance of GlcNAc-mediated signaling is made clear by disruption of OGT, which leads to embryonic lethality in mice (11). GlcNAcylated proteins have been implicated in several diseases including cancer and neurological conditions such as Alzheimer's disease (7). In addition, many groups have studied the role of GlcNAc in diabetes due to the direct correlation between increased protein GlcNAcylation and insulin resistance (4).

Previous studies have determined that many GlcNAcylated proteins are also phosphoproteins. Further investigation revealed that both modifications were remarkably similar in regards to modification reversibility, presence on both cytoplasmic and nuclear proteins, substoichiometric abundance and rapid changeover following cell activation (3, 4, 12). Consequently, a hypothetical model emerged that suggested GlcNAc and phosphate are involved in an interplay that serves to not only increase cellular diversity, but also to regulate GlcNAc- or phosphate-mediated signaling responses (4, 12). Crosstalk between GlcNAcylation and phosphate is possible because Ser and Thr residues can be modified with either moiety. This supports a model where GlcNAc can act as a competitive or reciprocal modification to phosphorylation by either identical or nearby site occupancy (2, 4, 13). Further evidence for interplay between phosphorylation and GlcNAcylation resulted from the discovery that OGT associates with protein phosphatase 1 to form an active complex capable of both dephosphorylation and GlcNAcylation (14).

Even though *O*-GlcNAc was discovered nearly 30 years ago and many GlcNAcylated proteins play a role in several diseases, the number of GlcNAc site identifications is low compared to known phosphorylation sites. By 2008, less than 15% of the known 600 GlcNAcylated proteins were site mapped (4). This is mainly attributed to several difficulties of GlcNAc characterization using conventional mass spectrometry. First, detection by MS can be quite difficult especially for highly complex peptide samples because GlcNAcylation is present in substoichiometric abundances (**Figure 3.2A**). Although a few GlcNAc-specific enrichment techniques

have been reported, there is not a widely accepted, efficient or reproducible method for routine enrichment (15-19). Second, RP HPLC separations do not adequately separate substoichiometric GlcNAcylated peptides from their unmodified counterparts (**Figure 3.2A**). This can lead to severe ion suppression, which precludes detection by MS. Third, GlcNAc is extremely labile in the gas phase under CAD conditions, which prevents site localization upon MS/MS due to the lack of GlcNAc-containing fragment ions (**Figure 3.2B**).

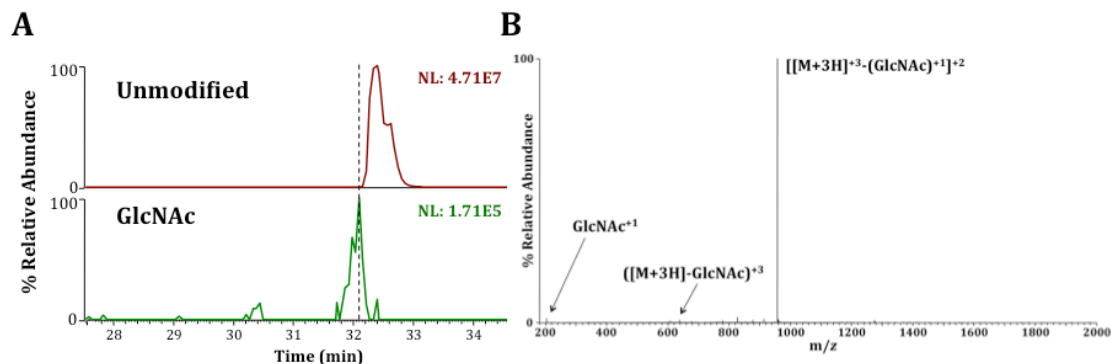


Figure 3.2 GlcNAcylated peptide behavior under C18 RPLC-MS/MS. (A) GlcNAcylated peptides are commonly substoichiometric and poorly separated from the unmodified counterpart peptide using RP-HPLC, which leads to severe ion suppression. (B) CAD MS/MS of a GlcNAcylated peptide results in essentially three ions that correspond to loss of GlcNAc from the precursor ion, without GlcNAc-containing fragment ions.

In order to circumvent many issues associated with MS-based GlcNAc characterization, the Hunt laboratory has developed methods that utilize immunoprecipitated proteins and incorporates ETD for site identification (18, 20, 21). Selective immunoprecipitation increases the dynamic range of MS experiments and improves the probability of GlcNAc detection by limiting sample complexity to a single protein. ETD is ideal for determining both GlcNAc and phosphorylation sites because it does not require vibrational excitation for backbone cleavage, which can

cause losses of GlcNAc and phosphate moieties from modified peptides. Efficient ETD fragmentation requires peptide charge states greater than +2. However, observation of many GlcNAcylated peptides has shown that modified residues are often located in regions rich in Ser and Thr with very few basic residues (Arg and Lys). The Hunt laboratory has optimized methods that facilitate peptide charge enhancement in order to avoid this issue. This chapter describes efforts aimed at developing and applying these mass spectrometric techniques for identifying *O*-GlcNAc post-translational modifications on human homeodomain-interacting protein kinase 1 (HIPK1).

3.2.2 HIPK1

Homeobox genes control organogenesis, determine cell fate and encode homeoproteins that specifically bind DNA sequences to regulate gene transcription (22). Homeoproteins encoded by the NK-2 homeobox gene family have important functions in embryonic development and mutations within this gene have been implicated in human congenital heart disease (23, 24). In order to determine proteins that interact with NK homeoprotein transcription factors, a two-hybrid screen used a portion of the Nkx1.2 homeoprotein as bait. As a result, a new class of nuclear kinases was discovered that act as NK homeoprotein transcriptional co-repressors. This group of kinases was later named the homeodomain-interacting protein kinases (HIPKs) (25).

The HIPK family contains four members: HIPK1, HIPK2, HIPK3 and HIPK4.

HIPKs 1-3 display high overall structural identity whereas HIPK4 is similar only in the kinase domain (26). Since their initial discovery, HIPKs have been found to localize to both the nucleus and cytoplasm and participate in a large variety of functions including signal transduction and the regulation of transcription and histone PTM (27-30). Mice deficient in HIPK1 or HIPK2 are seemingly healthy whereas HIPK1/2 double knockouts are embryonic lethal. This suggests that HIPK1 and HIPK2 likely share some functional responsibilities. Of the four HIPKs, HIPK2 is the most studied while HIPK1 studies are few (31).

HIPK1 is a 130 kDa kinase that shares distinct structural domains with HIPK2/3 that exhibit high sequence identity (**Figure 3.3**) (31). A functional Ser/Thr kinase domain is located at the extreme N-terminus. Interactions with homeodomain DNA sequences occur through the central region. PEST and YH amino acid-rich domains are located near the C-terminus.

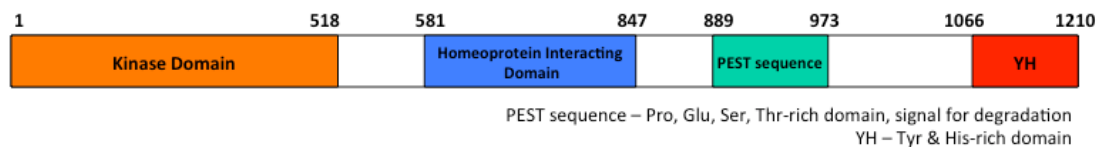


Figure 3.3 Structural domains of human HIPK1. HIPK1 likely binds to DNA through the homeodomain-interacting domain while interactions with AIP1 and ASK1 proteins are known to occur in the C-terminal YH domain.

One known function of HIPK1 involves pro-apoptotic signaling. In particular, HIPK1 has been implicated in the Fas-mediated apoptosis signal-regulating kinase 1 (ASK1)-JNK/p38 pathway that is activated by tumor necrosis factor (TNF) α and

reactive oxygen species (31-33). Translocation of HIPK1 from the nucleus to the cytoplasm is critical for this process. Initially, nuclear HIPK1 is modified by small ubiquitin-like modifier (SUMO). Upon TNF α stimulation, HIPK1 is desumoylated by SENP1. This facilitates translocation to the cytoplasm where HIPK1 associates with the ASK1-interacting protein (AIP1)-ASK1 complex. After HIPK1 binds, ASK1 dissociates from inhibitory proteins, thioredoxin (Trx) and 14-3-3. Removal of these proteins promotes activation of ASK1 to result in ASK1-JNK/p38 activation and ultimately, apoptosis (**Figure 3.4**) (32, 34).

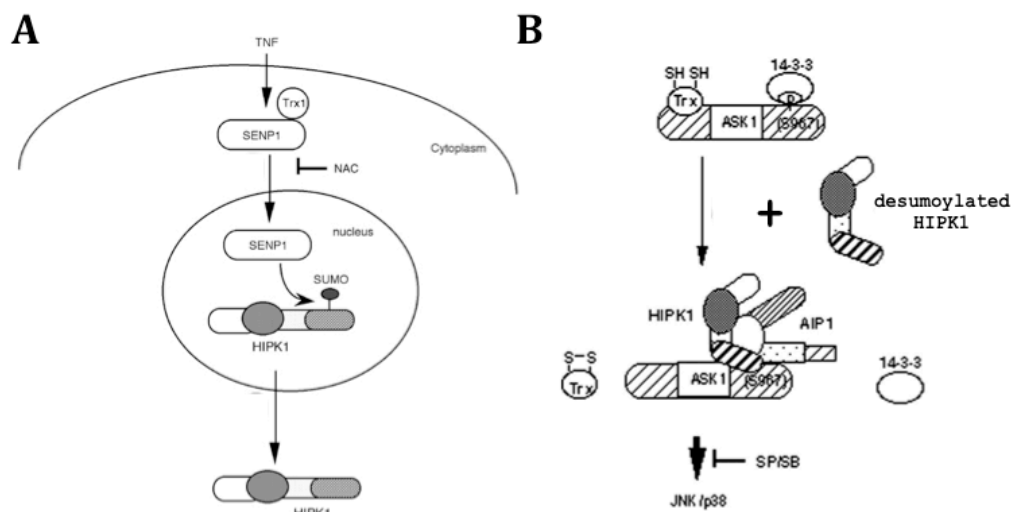


Figure 3.4 TNF α -induced HIPK1 desumoylation by SENP1 activates ASK1-JNK signaling to promote apoptosis. (A) SENP1 desumoylates HIPK1 in the nucleus to result in translocation to the cytoplasm. (B) Desumoylated HIPK1 and AIP1 bind the inactive ASK1-Trx-14-3-3 complex, resulting in dissociation of the ASK1 inhibitors Trx and 14-3-3. An active HIPK1-AIP1-ASK1 complex results and allows activation of JNK/p38 signaling and apoptosis. HIPK1 binds AIP1/ASK1 through its YH domain. Figures adapted with permission from (32, 34).

HIPK1 sumoylation is suspected to localize to Lys side chains at one of several aKXE consensus sequences in one of the following residues, where 'a' is any aliphatic residue and 'X' represents any amino acid. The Lys residues contained within these sequences are K25, K440, K556, K1203 (32). Only two other PTMs

have been identified on HIPK1 thus far; phosphorylation sites Y352 and S1200 were identified by high throughput MS analyses, but their functions are unknown (35-37).

In general, most protein GlcNAcylation has been shown to correlate with cellular UDP-GlcNAc concentrations. However, GlcNAcylation on some proteins is independent of cellular UDP-GlcNAc concentration, which suggests that GlcNAcylation must also be controlled by a separate mechanism (38). Cheung et al hypothesize that specific protein interactions account for a portion of OGT's specificity. This is in agreement the observation that discovered proteins which interact with OGT via an N-terminal tetratricopeptide repeat domain at the OGT are also GlcNAcylated (39). HIPK1 was recently identified as an OGT-interacting protein using a yeast two-hybrid screen (38). A HIPK1/OGT interaction is particularly interesting due the known interplay between phosphorylation and *O*-GlcNAcylation and the possibility that *O*-GlcNAcylation could serve to regulate apoptosis. As a result, the Hart laboratory began to study HIPK1 GlcNAcylation to determine possible regulatory roles in HIPK1 translocation and apoptosis. To identify HIPK1 *O*-GlcNAc sites, we applied sensitive LC-MS/MS techniques that promote simultaneous characterization of *O*-GlcNAc and phosphate modifications.

3.3 Materials

3.3.1 Reagents

Gemini Bio-Products (West Sacramento, CA)

Fetal bovine serum

Invitrogen (Carlsbad, CA)

β -1,4-galactosyltransferase

Lipofectamine 2000

Uridine diphosphate-N-galactose (UDP-Gal)

Mallinckrodt, Inc. (Paris, KY)

Acetonitrile, HPLC grade, $\geq 99.8\%$ purity

Mediatech, Inc. (Manassas, VA)

Dulbecco's modified Eagle's medium

Pierce Protein Research Products (Thermo Scientific, Rockford, IL)

CTD 110.6 (anti-*O*-GlcNAc) antibody

PQ Corporation (Valley Forge, PA)

Kasil® 1624

Lithisil®

Promega Corporation (Madison, WI)

Trypsin, sequencing grade, modified

Roche (Penzberg, Germany)

Endoproteinase Asp-N, sequencing grade

Chymotrypsin, sequencing grade

Sigma Aldrich (St. Louis, MO)

Ammonium bicarbonate

Angiotensin I acetate salt hydrate, $\geq 90\%$ purity (human)

Azulene

Dithiothreitol

EDC, (N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride)

Flag-M2 (anti-Flag) antibody

Formamide

Glacial acetic acid, $\geq 99.9\%$ purity

Iodoacetamide

Pyridine

Tetramethylammonium silicate

Vasoactive intestinal peptide fragment 1-12, $\geq 97\%$ purity (human, porcine, rat)

3.3.2 Equipment & Instrumentation

Agilent Technologies (Palo Alto, CA)

HP1100 high performance liquid chromatograph

Labconco Corporation (Kansas City, MO)

Centrivap centrifugal concentrator

Polymicro Technologies, LLC (Pheonix, AZ)

360 μm o.d. x 50 μm i.d. polyimide coated fused silica capillary

360 μm o.d. x 75 μm i.d. polyimide coated fused silica capillary

360 μm o.d. x 150 μm i.d. polyimide coated fused silica capillary

Sutter Instrument Co. (Navato, CA)

P-2000 microcapillary laser puller

Thermo Scientific (San Jose, CA/Bremen, Germany)

LTQ-Orbitrap hybrid mass spectrometer

LTQ-FT-ICR hybrid mass spectrometer

YMC Co., Ltd. (Kyoto, Japan)

C18 resin (5-20 μm diameter, 300 \AA pore size)

C18 resin (5 μm diameter, 120 \AA pore size)

Zeus Industrial Products (Orangeburg, SC)

0.060 in o.d. x 0.012 in i.d. Teflon tubing

3.4 Methods

The sample preparation and cell biology-based experiments were completed by Dr. Ronald Copeland in Gerald Hart's laboratory in the Department of Biological Chemistry at Johns Hopkins University.

3.4.1 Cell Culture & Flag-HIPK1 Transfection

HEK293 cells were grown in Dulbecco's modified Eagle's medium with 25 mM glucose containing 10% (v/v) fetal bovine serum. The human wild-type Flag-HIPK1 (pFlag-CMV2, NotI) gene construct was a gift from Dr. Yoshihiko Miyata (Kyoto University). HEK293 cells were transfected with 1 μg plasmid DNA using Lipofectamine 2000 according to manufacturer's instruction.

3.4.2 Affinity Purification of Flag-tagged HIPK1

HEK293 cells were harvested 48 h after transfection and lysed in 50 mM Tris lysis buffer (pH 7.4) that contained 2% Triton-X 100 and 1X protease and phosphatase inhibitors. The lysate was incubated on ice for 30 m, sonicated and centrifuged to remove cell debris. A 50% slurry of Flag-agarose beads were washed with cold PBS prior to incubation with 2 mg of HEK293 lysate overnight at 4°C.

Anti-Flag agarose beads were subjected to centrifugation at 8000 rpm for 2 min after which the supernatant was discarded and the beads were washed with 1 mL of 50 mM Tris buffer (pH 7.4) without protease inhibitors.

3.4.3 Immunoblot and Autoradiography Analysis of O-GlcNAcylated HIPK1

To confirm that expressed Flag-HIPK1 was GlcNAcylated *in vivo*, HIPK1 and pFlag-CMV2 (empty vector control) were transfected into HEK293 cells with Lipofectamine for 48 h. Pre-cleared lysates were immunoprecipitated with anti-Flag agarose beads, subjected to SDS-PAGE analysis and transferred to nitrocellulose for Western blotting using Flag-M2 and CTD110.6 antibodies.

For autoradiography analysis, pre-cleared lysates were prepared as above and treated with ^3H -UDP-Gal and β -1,4-galactosyltransferase overnight at 4°C. Following incubation, immunoprecipitates were washed with RIPA buffer and subjected to SDS-PAGE analysis.

3.4.4 Proteolytic Digestions and Sample Preparation for Mass Analysis

Prior to proteolytic digestion, *in silico* digestions were performed on the human HIPK1 sequence (UniProt Accession Q86Z02) using ExPASy PeptideMass digestion simulator at <http://expasy.org/tools/peptide-mass.html>.

Flag-HIPK1 retained on beads was suspended in 50 mM Tris buffer (pH 7.4) and washed with three 50 μL aliquots of 100 mM ammonium bicarbonate (pH 8.0) to equilibrate the pH. The agarose beads were resuspended in 30 μL of 100 mM

ammonium bicarbonate. Cys residues were reduced using 10 μ L of 0.340 mM DTT in 100 mM ammonium bicarbonate for 1.5 h at room temperature. Directly afterwards, Cys residues were carbamidomethylated using 10 μ L of 0.680 mM iodoacetamide in 100 mM ammonium bicarbonate for 45 min in the dark at room temperature. On beads proteolysis proceeded with endoproteinase Asp-N at room temperature for 7 h using 125 ng of enzyme. The supernatant was removed, and quenched using 2 μ L of glacial acetic acid prior to storage at -35°C .

A trypsin sub-digestion was later performed. A 20% aliquot of the Asp-N digestion was brought to pH 8 using 1 mM ammonium bicarbonate. Tryptic sub-digestion proceeded for 6 h at 37°C using 25 ng enzyme. The digestion was quenched using 3 μ L glacial acetic acid and stored at -35°C until mass analysis.

A second sub-digestion was performed using chymotrypsin. The remaining Asp-N digestion was dried down using the Centrivap concentrator and resuspended in 30 μ L of 100 mM ammonium bicarbonate. The sub-digestion proceeded for 5 h at 37°C using 75 ng enzyme and was quenched using 3 μ L glacial acetic acid prior to storage at -35°C .

3.4.5 HPLC Capillary Column Assembly

An 8 cm precolumn was constructed using 360 μm o.d. x 75 μm i.d. fused silica capillary with a 2 mm Kasil® 1624 frit and packed with 5 cm of irregular C18 reverse phase resin (5-20 μm diameter, 300 Å pore size). A 9 cm analytical column

was constructed using 360 μm o.d. x 50 μm i.d. fused silica capillary containing a 1 mm Lithisil® frit. The analytical column was packed with 8 cm of regular C18 resin (5 μm diameter, 120 Å pore size) and equipped with a laser-pulled, integrated electrospray emitter tip as in (40). Both columns were prepped using several picomoles of angiotensin to optimize chromatographical separation.

A 12 cm “clean up” column was constructed using 360 μm o.d. x 75 μm i.d. fused silica capillary with a 2 mm Kasil® 1624 frit and packed with 10 cm of irregular C18 reverse phase resin (5-20 μm diameter, 300 Å pore size). Prior to use, the column was prepped using 30 picomoles of angiotensin.

3.4.6 HPLC-MS/MS Analysis

For mass spectrometric analysis of the endoproteinase Asp-N digestion, a 5% and a 15% aliquot were used for data dependent CAD and ETD analyses, respectively. The trypsin sub-digestion was split in half for separate CAD and ETD analyses. One-sixth of the chymotrypsin sub-digestion was analyzed using data dependent CAD. For each analysis, peptides were pressure loaded onto the precolumn at a flow rate 1 $\mu\text{L}/\text{min}$. Afterwards, known concentrations of internal standards angio and vaso were loaded and the precolumn was washed with 0.1 M acetic acid for 10 min at similar flow rates.

For mass analysis of each digestion, proteolytic peptides were gradient eluted at a split flow-generated rate of 60 nL/min into the mass spectrometer using

an Agilent 1100 chromatograph and linear LC gradient as follows: 0-60% Solvent B in 60 min, 60-100% Solvent B in 5 min, 100% Solvent B hold for 5 min, 100% - 0% B in 5 min, 100% Solvent A for 20 min (Solvent A: 0.1 M acetic acid in water and Solvent B: 70% acetonitrile in 0.1 M acetic acid in water).

Mass analyses incorporated an initial high resolution (60,000 at m/z 400) MS1 survey scan in the Orbitrap or FT-ICR cell that occurred simultaneously with low resolution, data dependent MS/MS scans in the linear ion trap. CAD analyses incorporated 10 data dependent CAD MS/MS scans. ETD analyses incorporated 6 data dependent ETD MS/MS scans. Data dependence parameters for CAD MS/MS included charge state rejection of +1 species, monoisotopic precursor selection “on,” repeat count of 2, repeat duration of 30 s, exclusion list duration of 20 s, and precursor isolation window of 3 m/z . ETD parameters included 50-75 ms reaction time, charge state rejection “on” for $z = +1$ and $+2$, FTMS AGC target of $2e5$, ITMSⁿ AGC target of $1e4$, and azulene ETD reagent target of $2e5$.

All data were searched using OMSSA (version 2.1.1) using MS/MS peak lists generated by Bioworks Browser (version 3.3.1 SP1) against a generated human Flag-HIPK1 sequence Blast database. Searches of CAD and ETD data shared the following parameters: Asp-N or no enzyme specificity for sub-digestions, E-value cutoff of “1,” missed cleavage count of 3, and variable modifications included carbamidomethylation of Cys, oxidation of Met, *O*-GlcNAcylation of Ser and Thr, phosphorylation of Ser, Thr and Tyr. Mass tolerances were set at ± 0.01 Da and

± 0.35 Da for precursor and product ion masses, respectively. OMSSA provided the removal of reduced charge species in ETD spectra. All pertinent spectra were manually confirmed to ensure confident peptide and PTM identification.

3.4.7 Histamine Derivatization of HIPK1

Peptides generated from the Asp-N/chymotryptic digestion were subjected to charge state modification by histamine derivatization. Prior to derivatization, peptides were purified from endoproteinases. The entire aliquot was loaded onto a 360 μm o.d. x 150 μm i.d. "clean up" column and rinsed with 0.1 M acetic acid for 10 min. Peptides were gradient eluted into a clean Eppendorf tube from 0-70 min using the following gradient: 0-80% Solvent B in 40 min, 80% Solvent B hold for 30 min, 80%-100% Solvent B in 10 min, 100%-0% Solvent B in 5 min.

The purified peptides were dried using the Centrivap concentrator and resuspended in 20 μL of 1 M histamine in pyridine buffer (1 M pyridine HCl, pH 5-5.5). A 5 μL aliquot of 0.1 M EDC in pyridine buffer was added and the reaction mixture was sonicated for at room temperature for 2 h in the dark. The derivatized peptides were dried and resuspended in 60 μL 0.1% acetic acid (pH 3). These peptides were split in half and analyzed using data dependent CAD or ETD analysis with methods previously described with slight modification: two peptides were subjected to targeted ETD MS/MS analysis in the +3 charge state (GlcNAcylated VHQPVS VGPSLL and phosphorylated TFQHG SPLHSTGH PH). OMSSA parameters

were similar to those previously used and included variable histamine modification (+93 Da) on Asp, Glu, and C-terminal residues.

3.5 Results & Discussion

The Flag-HIPK1 sequence transfected into HEK293 cells for purification is listed in **Figure 3.5**. This was used to generate databases for OMSSA searches prior to manual data analysis.

N-terminal FLAG tag: MDYKDDDDKLA

```

1 MASQLQVFSP PSVSSSAFCS AKKLEIEPSG WDVSGQSSND KYITHSKTLP
51 ATQGGANSSH QVANFNIPAY DQGLLLPAPA VEHIVVTAAD SSGSAATSTF
101 QSSQTLTHRS NVSLLEPYQK CGLKRKSEEV DSNGSVQIEE EHPPLMLQNR
151 TVVGAAATTT TVTTKSSSSS GEGDYQLVQH EILCSMTNSY EVLEFLGRGT
201 FGQVAKCWRK STKEIVAIKI LKNHPSYARQ GQIEVSILSR LSSENADENY
251 FVRSYECFQH KNHTCLVFEM LEONLYDFLK QNKFSPLPLK YIRPILQQVA
301 TALMKLKSLG LIHADLKPEN IMLVDPVRQP YRVKVIDFGS ASHVSKAVCS
351 TYLQSRYYRA PEIILGLPFC EAIDMWSLGC VIAELFLGWP LYPGASEYDQ
401 IRYISQTQGL PAEYLLSAGT KTRFFNRDP NLGYPLWRLK TPEEHELETG
451 IKSKEARKYI FNCLDDMAQV NMSTDLEGTD MLAEKADRRE YIDLLKMLT
501 IDADKRITPL KTLNHQFVTM THLLDFPHSN HVKSCFQNME ICKRRVHMYD
551 TVSQIKSPFT THVAPNTSTN LTMSFSNQLN TVHNQASVLA SSSTAAAATL
601 SLANSDVSLN NYQSALYPSS AAPVPGVAQQ GVSLLQPGTTQ ICTQTDPFQQ
651 TFIVCPPAFQ TGLQATTKHS GFPVRMDNAV PIVPQAPAAQ PLQIQSGVLT
701 QGSCPTLMVA TLHPQVATIT PQYAVPFTLS CAAGR PALVE QTA AVLQAWP
751 GGTQQILLPS TWQQLPGVAL HNSVQPTAMI PEAMGSGQQQ ADWRNAHSHG
801 NQYSTIMQQP SLLTNHVTLA TAQPLNVGVA HVVRQQQSSS LPSKKNKQSA
851 PVSSKSSLDV LPSQVYSLVG SSPLRTTSSY NSLVPVQDQH QPIIIPDTPS
901 PPSVITIRS DTDEEDNKY KPSSSGLKPR SNVISYVTVN DSPSDSSSL
951 SPYSTDTLSA LRGNSGSVLE GPGRVVADGT GTRTIIIVPPL KTQLGDCTVA
1001 TQASGLLSNK TKPVASVSGQ SSGCCITPTG YRAQRGGTSA AQPLNLSQNG
1051 QSSAAPTSEQ RSSNPAPRRQ QAFVAPLSQA PYTFQHGSPH HSTGPHLAP
1101 APAHLPSQAH LYTYAAPTSA AALGSTSSIA HLFSPQSSR HAAAYTTHPS
1151 TLVHQVPVSV GPSLLTSASV APAQYQHQFA TQSYIGSSRG STIYTGYPPLS
1201 PTKISQYSYL

```

Figure 3.5 Amino acid sequence of Flag-tagged human HIPK1.

3.5.1 Affinity Purification of Flag-tagged HIPK1

Anti-Flag purification of Flag-HIPK1 resulted in Coomassie-stainable amounts of protein for MS analysis (**Figure 3.6**). Flag-based IPs are efficient for

tagged protein purification, but can also result in the simultaneous non-specific purification of proteins without a Flag tag. As a result, multiple bands are evident in the HIPK1 lane.

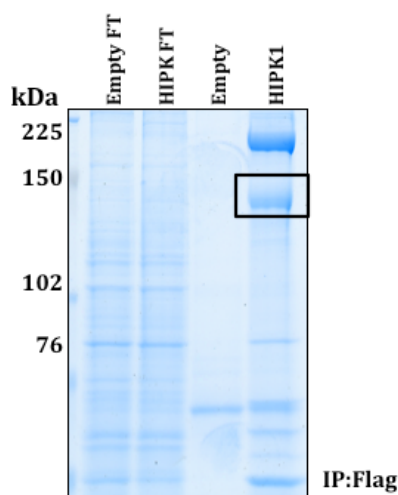


Figure 3.6 Immunoprecipitation of Flag-HIPK1. Immunoprecipitates and flow through (FT) aliquots were subjected to SDS-PAGE and proteins were visualized by Coomassie stain. The protein band at 130 kDa corresponds to Flag-HIPK1.

3.5.2 Immunoblot and Autoradiography Analysis of *O*-GlcNAcylated HIPK1

The yeast two-hybrid screen suggested a HIPK1-OGT interaction, based on which HIPK1 could be *O*-GlcNAcylated. Two preliminary experiments supported this conclusion. Western blotting using the pan *O*-GlcNAc antibody CTD110.6 resulted in a distinct band for a GlcNAcylated protein that exhibited the same electrophoretic mobility as Flag-HIPK (**Figure 3.7A**).

Protein GlcNAcylation can be detected using GalT-catalyzed addition of UDP-galactose to GlcNAc residues. This reaction requires the Y289L mutant form of GalT (41) that covalently attaches tritium-labeled galactose (^3H -UDP-Gal) to terminal GlcNAc moieties only. Upon analysis of Flag-HIPK1, autoradiography displays an

intense band near ~130 kDa, which correlates to Flag-HIPK1 (**Figure 3.7B**).

Blotting using anti-Flag antibody reveals a band at the same region. These two experiments suggest that Flag-HIPK1 is indeed modified by *O*-GlcNAc. The series of contaminating proteins evident in the anti-Flag Western blot demonstrates again that many proteins are non-specifically recognized by the anti-Flag epitope, although only one protein is heavily GlcNAcylated.

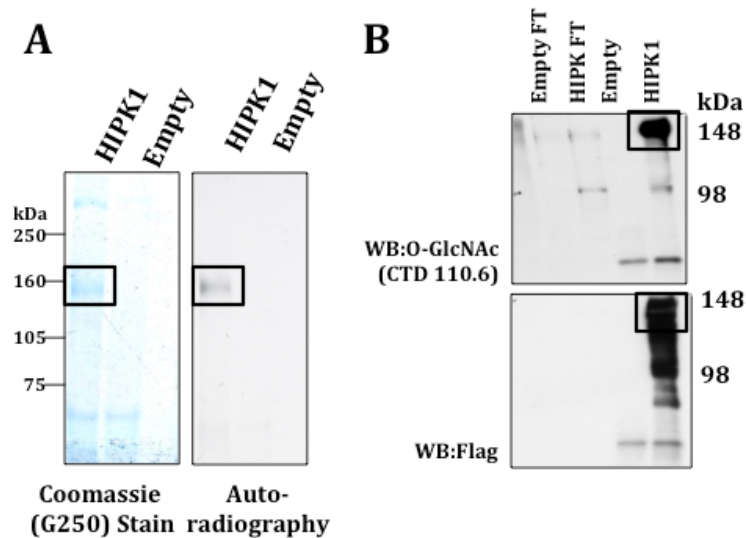


Figure 3.7 Flag-HIPK1 is GlcNAcylated *in vivo*. (A) ^3H -UDP-Gal is enzymatically incorporated into Flag-HIPK1 by incubation with GalT. (B) Western blotting using the pan anti-GlcNAc antibody CTD110.6 demonstrates that Flag-purified HIPK1 is GlcNAcylated.

3.5.3 *In silico* Proteolytic Digestion

In silico digestions of Flag-HIPK1 with commonly used enzymes (trypsin, endoproteinases Asp-N, Glu-C, Lys-C, chymotrypsin) or chemical cleavage by cyanogen bromide demonstrated that proteolysis by a single method was not optimal. Digestion using endoproteinase Asp-N provided the best overall

proteolysis for LC-MS/MS analysis by yielding detectable peptides that include nearly all *O*-GlcNAc modifiable (Ser, Thr) and phosphorylatable residues (Ser, Thr, Tyr) (**Table 3.1**). Only a few Asp-N-generated peptides were small enough to escape detection using C18 RP-HPLC. Of these peptides, only a few contained Ser, Thr or Tyr residues. However, some Asp-N generated peptides were so large (>100 residues) that they were unlikely to elute from the C18 RP-HPLC column. These peptides made it clear that sub-digestions were necessary.

Mass (Da)	Position	Peptide Sequence	#STY	Comments
3297.6646	1-31	MASQLQVFSPPSVSSSAFCSAKKLEKIEPSGW	8	+
793.3322	32-39	DVSGQSSN	3	+
3451.6665	40-70	DKYYTHSKTLPATQGGANSSHQVANFNIPAY	9	+
1914.0694	71-89	DQGLLLPAPAVEHIVVTAA	1	+
4457.2001	90-130	DSSGSAATSTFQSSQTLTHRSNVSLLEPYQKCGLKRKSEEV	14	+
4358.1416	131-173	DSNGVQIIEHPPLMLQNRTVVGAAATTTT VTTKSSSSSGEG	14	+
8310.3052	174-246	DYQLVQHEILCSMTNSYEVLEFLGRGTFGQVAKCWKRSTKEIVAIIKLNHPSYARQQIEVSILSRLSSENA	14	?
3799.6978	247-276	DEYNFVRSYECFQHKNGTCLVFEMLEQNLV	5	+
4333.5348	277-314	DFLKQNKFSPLPLKYIRPILQQVATALMKLKSGLIHA	4	+
1171.6391	315-324	DLKPENIMLV	0	+
1469.8587	325-336	DLPVRQPYRVKVI	1	+
4092.0357	337-373	DFGSASHVSKAVCSTYLQSRYYRAPEIILGLPFCEAI	9	+
2817.3302	374-398	DMWSLGCVIAELFLGWPLYPGASEY	4	+
3474.8128	399-428	DQIRYISQTQGLPAEYLLSAGTKTTRFFNR	8	+
4288.2222	429-464	DPNLGYPLWRLKTPEEHELETGIKSKEARKYIFNCL	5	+
134.0448	465	D	0	-
996.4125	466-474	DMAQVNMST	2	+
534.2406	475-479	DLEGT	1	?
777.3811	480-486	DMLAEKA	0	+
851.4370	487-492	DRREYI	1	?
1074.6591	493-501	DLLKKMLTI	1	+
205.0819	502-503	DA	0	-
2506.3962	504-524	DKRITPLKTLNHQFVTMTHLL	4	+

3106.4379	525-549	DFPHSNHVKSCFQNMEICKRRVHMY	3	+
5733.8300	550-605	DTVSQIKSPFTTHVAPNTSTNLTMSFSNQL NTVHNQASVLAASSTAAAATLSLANS	20	+
4090.0437	606-645	DVSLLNYSALYPSSAAPVPGVAQQGVSLQ PGTTQICTQT	11	+
3449.7133	646-676	DPFQQTIVCPPAFQTGLQATTKHSGFPVR M	5	+
11849.1255	677-791	DNAVPIVPQAPAAQPLQIQSGVLTQGSCTPL MVATLHPQVATITPQYAVPFTLSAAGRPA LVEQTA AVLQAWPGGTQQILLPSTWQQLP GVALHNSVQPTAMIPEAMGSGQQLA	17	-
7273.7424	792-858	DWRNAHSHGNQYSTIMQQPSLLTNHTVTL ATAQPLNVGVAVHVVRRQQSSSLPSKKNKQS APVSSKSSL	17	?
3093.6102	859-887	DVLPSQVYSLVGSPLRRTSSYNLVPVQ	11	+
1060.5785	888-896	DQHQPPIIP	0	+
1468.8005	897-910	DTPSPVSVITIRS	5	+
235.0924	911-912	DT	1	+
521.1725	913-916	DEEE	0	-
2653.3944	917-940	DNKYKPSSSGLKPRSNVISYVTVN	8	+
318.1296	941-943	DSP	1	-
221.0768	944-945	DS	1	-
1043.4527	946-955	DSSLSPYST	7	+
2155.1465	956-977	DTLSALRGNLGSVLEGPRVVA	4	+
1867.0647	978-995	DGTGTRTIIVPLKTLQLG	4	+
22273.0414	996-1210	DCTVATQASGLLSNKTTPVASVSGQSSGCCI TPTGYRAQRGGTSAAPLNLSQNNQSSAAP TSQERSSNPAPRRQAFVAPLSQAPYTFQH GSPLHSTGHPHLAPAPAHLPQAHLYTYAA PTSAAALGSTSSIAHLFSPQSSRHAAAYTT HPSTLVHQVPVSVGPSLLTSASVAPAQYQH QFATQSYIGSSRGSTIYTGYPSPKISQYSYL	67	-

Table 3.1 *In silico* digestion of human HIPK1 with endoproteinase Asp-N. Predicted unmodified masses, residue positions and the number of STY residues per peptide are listed. Peptides expected to be detected are denoted with a "+" in Comments; questionable peptides are denoted with a "?"; peptides not expected to be detected are denoted by "-". Peptides denoted by "+" constitute < 60% of the total sequence.

3.5.4 CAD and ETD Analysis of Asp-N HIPK1 Peptides

Initially, a 5% aliquot of the total Asp-N digest was analyzed using data dependent CAD fragmentation. This provided a cursory view of sample purity and peptide concentration. Following analysis, the sample was determined to contain a significant amount of polyethylene glycol (PEG) contamination, evident as ions separated by 44 Da in the +1 charge state. This contamination found in mass

spectrometric analyses is commonly observed in the Hunt laboratory when peptides

are stored at length in plastic tubes in basic solution (42, 43). While this situation is avoidable, a miscommunication resulted in sample storage in base and shipment from Johns Hopkins University in 50 mM Tris buffer (pH 7.4). Unfortunately, these polymeric species are retained by C18 RP material and elute on a similar timescale (20 min – 60 min) to proteolytic peptides (**Figure 3.8**). Despite this contamination, CAD analysis afforded the identification of several HIPK1 peptides in the base peak chromatogram. HIPK1 peptide signals were lower than those corresponding to 250 fmol of angio and vaso. In order to increase the amount of peptide analyzed and resulting MS/MS spectral quality, a 15% fraction was subsequently analyzed with ETD to provide a complementary MS/MS data set. This yielded HIPK1 peptide abundances near 500 fmol for the most abundant HIPK1 peptides (**Figure 3.8**). The methods used for relative abundance calculations are outlined in Chapter 1. Since both Asp-N mass analyses demonstrated that HIPK1 peptides were more abundant than those from other proteins, it was clear that the IP was effective for efficient HIPK1 purification.

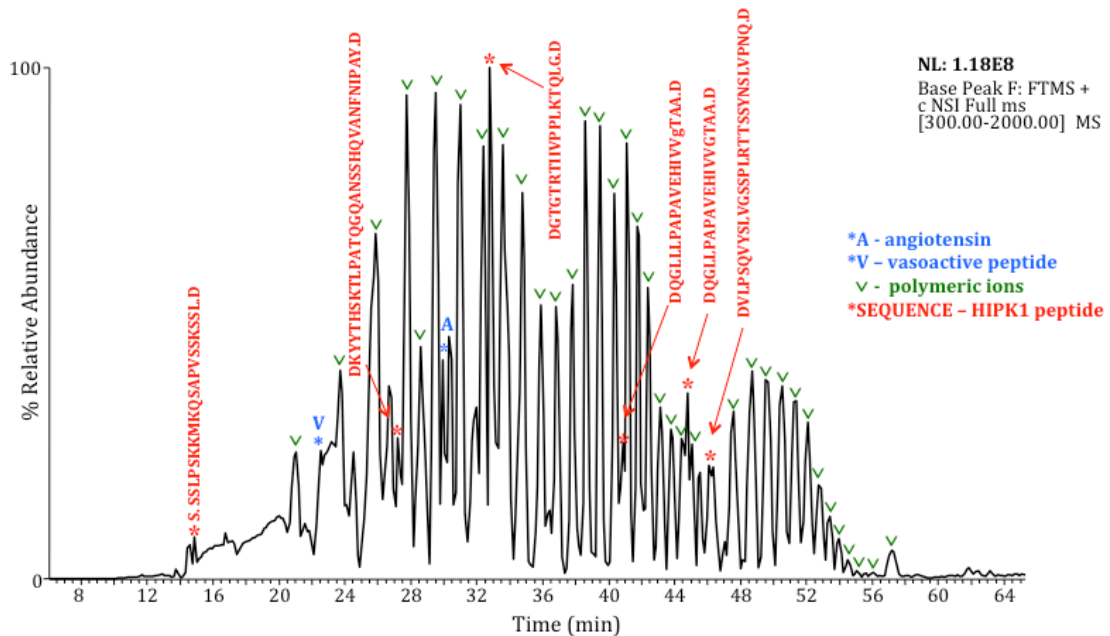


Figure 3.8 Base peak chromatogram of endoproteinase Asp-N-generated HIPK1 peptides. LC-MS/MS analysis of a 5% aliquot of the total digest showed that many peaks in the base peak chromatogram were either HIPK1 proteolytic peptides or polymeric contaminants.

HIPK1 peptides were verified by manual inspection using both CAD and ETD data sets. These results were used to calculate coverage of the entire HIPK1 sequence. Upon compilation of CAD and ETD data, it was determined that Asp-N digestion alone resulted in 56% sequence coverage (**Figure 3.9**). Many peptides, particularly those near the C-terminus, were large and therefore unlikely to elute from C18 resin. From these results, it was clear that sub-digestion was required to perform a comprehensive PTM analysis of the entire HIPK1 sequence.

1 MASQLQVFSP PSVSSSAFCS AKKLKIEPSG WDVSGQSSND KYYTHSKTLP
 51 ATQGQANSSH QVANFNIPAY DQGLLLPAPA VEHIVVTAAD SSGSAATSTF
 101 QSSQTLTHRS NVSLLLEPYOK CGLKRKSEEV DSNGSVQIIE EHPPLMLQNR
 151 TVVGAATTT TVTTKSSSSS GEGDYQLVQH EILCSMTNSY EVLEFLGRGT
 201 FGQVAKCWKR STKEIVAIAKI LKNHPSYARQ GQIEVSILSR LSSENADDEYN
 251 FVRSYECFQH KNHTCLVFEM LEQONLYDFLK QNKFSPLPLK YIRPILQOVA
 301 TALMKLKSLG LIHADLKPEN IMLVDPVRQP YRVKVIDFGS ASHVSKAVCS
 351 TYLQSRYYRA PEIILGLPFC EAIDMWSLGC VIAELFLGWP LYPGASEYDQ
 401 IRYISQTOGL PAEYLLSAGT KTTRFFNRDP NLGYPLWRLK TPEEHELETG
 451 IKSKEARKYI FNCLDDMAQV NMSTDLEGTD MLAEKADRRE YIDLLKMLT
 501 IDADKRITPL KTLNHQFVTM THLLDFPHSN HVKSCFQNM ICKRRVHMYD
 551 TVSQIKSPFT THVAPNTSTN LTMSFSNQLN TVHNQASVLA SSSTAAATL
 601 SLANSDVSL NYQSALYPSS AAPVPGVAQQ GVSLOPGTTO ICTQTDPFQO
 651 TFIVCPPAFQ TGLQATTKHS GFPVRMNAV PIVPOAPAAQ PLQIQSGVLT
 701 QGSCPTLMVA TLHPQVATIT PQYAVPFTLS CAAGRPALVE QTAAVLQAWP
 751 GGTQQILLPS TWQQLPGVAL HNSVQPTAMI PEAMGSGQOL ADWRNAHSHG
 801 NOYSTIMQOP SLLTNHVTLA TAQPLNVGVA HVVRQQQSSS LPSKKNQOSA
 851 PVSSKSSLDV LPSQVYSLVG SSPLRTTSSY NSLVVPVQDQH QPIIIPDTPS
 901 PPVSVITIRS DTDEEEDNKY KPSSSGLKPR SNVISYVTVN DSPDSDSSLS
 951 SPYSTDTLSA LRGNSGSVLE GPGRVVADGT GTRTIIVPPL KTQLGDCTVA
 1001 TOASGLLSNK TKPVASVSGQ SSGCCITPTG YRAQRGGTSA AQPLNLSONQ
 1051 QSSAAPTQOE RSSNPAPRRO QAFVAPLSQA PYTFQHGSP HSTGHPHLAP
 1101 APAHLPSQAH LYTYAAPTSA AALGSTSSIA HLFSPQGSRR HAAAYTTHPS
 1151 TLVHQVPVSV GPSLLTSASV APAQYQHQFA TQSYIGSSRG STIYTGYP
 1201 PTKISQSYL

Figure 3.9 Proteolysis of HIPK1 using endoproteinase Asp-N results in 56% sequence coverage using CAD and ETD MS. Unambiguously identified residues are underlined. Some peptides vary from the normal Asp-N cleavage consensus sequence due to enzyme miscleavage.

Upon further inspection of the data, post-translationally modified peptides were detected. A precursor ion was identified that corresponded by mass (+1.4 ppm) to a singly GlcNAcylated HIPK1 peptide DQGLLLPAPAVEHIVVTA. Elution of this peptide occurred just prior to the previously confirmed unmodified counterpart peptide (**Figure 3.10A**). This characteristic elution profile results from a slight increase in hydrophilicity. CAD fragmentation confirmed the presence of GlcNAc due to three signature ions: 1) a GlcNAc⁺¹ oxonium ion (m/z 204.1), 2) neutral loss of GlcNAc from the precursor ion ([M+3H]-GlcNAc)⁺³, and 3) loss of GlcNAc⁺¹ ion to result in unmodified precursor in one charge state lower, ([M+3H]-GlcNAc⁺)⁺² (**Figure 3.10B**).

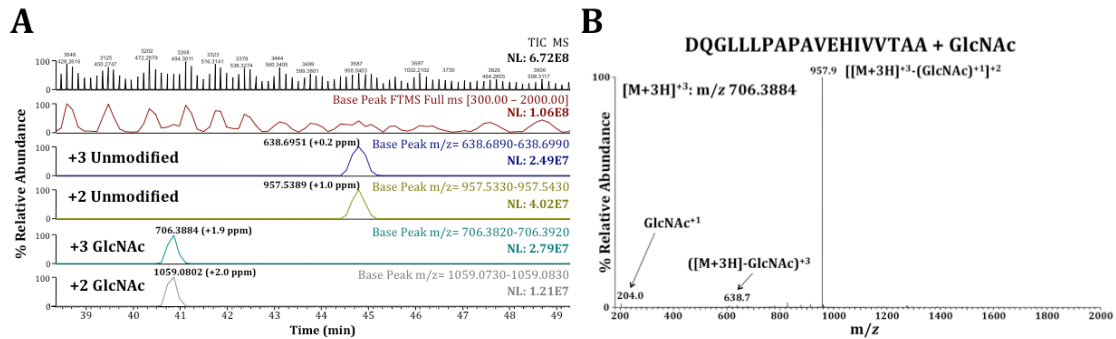


Figure 3.10 HIPK1 peptide DQGLLLPAPAVEHIVVTAA was found to be GlcNAcylated based on accurate mass and CAD MS/MS. (A) XICs of unmodified and GlcNAcylated HIPK1 peptide DQGLLLPAPAVEHIVVTAA demonstrate nearly equally abundant +2 and +3 charge states. (B) CAD MS/MS analysis of GlcNAcylated DQGLLLPAPAVEHIVVTAA yields virtually no fragment ions but instead displays three characteristic signature ions associated with loss of GlcNAc.

Although the three signature ions are sufficient for identifying peptides that are *O*-GlcNAc modified, they are useless for determining the site of GlcNAcylation. Localization cannot necessarily be assumed if a single Ser or Thr residue is present because upon natural endoglycosidase H activity, N-glycosylation is cleaved to a single GlcNAc on Asn residues (44). In order to be confident in *O*-GlcNAc sites, it is necessary to provide unequivocal evidence that unambiguously identifies the modified residue. ETD, which yields rapid peptide fragmentation free of vibrational excitation, can provide this data. ETD MS/MS analysis unambiguously identified the GlcNAc site to T87 (**Figure 3.11**). This example clearly demonstrates the utility of ETD for GlcNAc site identification.

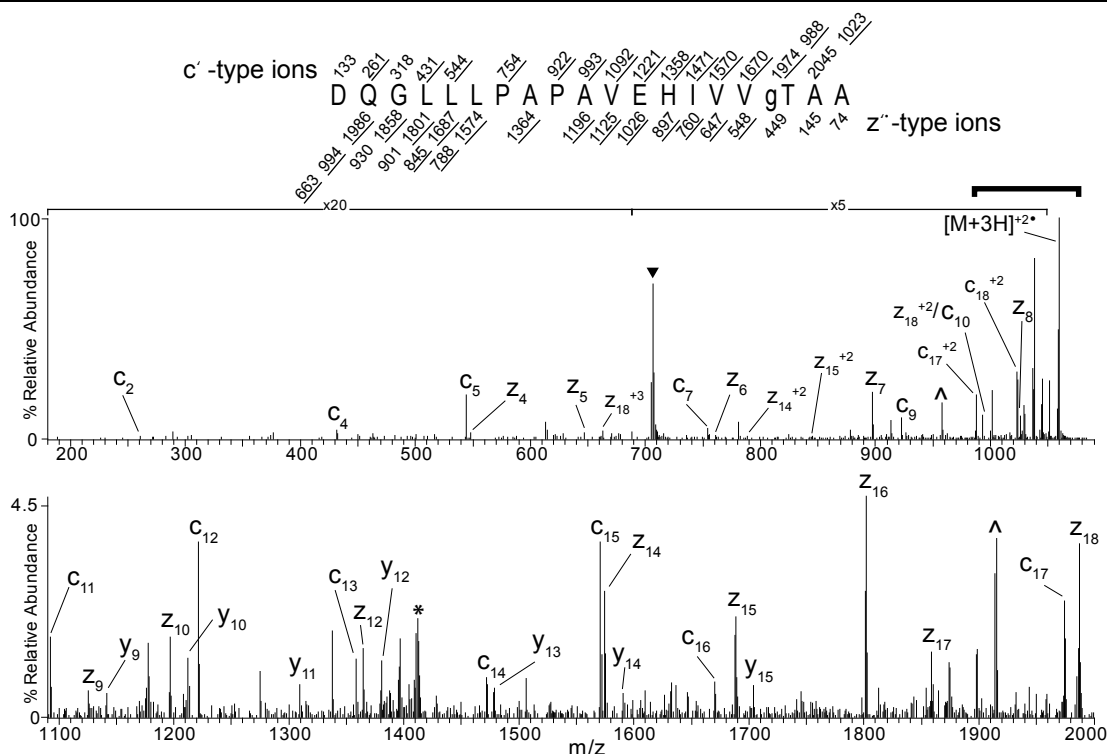


Figure 3.11 ETD MS/MS spectrum of singly GlcNAcylated peptide DQGLLLPAPAVEHIVVgTAA. Both c- and z⁺-type fragment ions detected are labeled within the spectrum and underlined in the peptide sequence. Neutral losses from fragment ions are designated by ^. Heavy bracketed ions correspond to neutral losses from labeled reduced charge species (RCS). RCS from species that coelute and fall within the precursor isolation window are labeled with *. The inverted triangle (▼) designates ions that fall within the precursor isolation window. Note that no loss of GlcNAc (204 m/z) is observed. Nearly all fragment ions expected are detected and demonstrate O-GlcNAcylation of T87.

In addition to the T87 GlcNAc site, several phosphorylation sites were validated using CAD and ETD MS/MS spectra (**Table 3.2**). CAD contributed fragment ions to assign phosphorylation sites on T897 and S899 while ETD provided localization of the remaining sites.

Site	Peptide	Modification
T87	DQGLLLPAPAVEHIVVgTAA	O-GlcNAc
S127	EPYQKcGLKRKRKpSEEV	Phosphate
T441	DPNLGYPLWRLKpTPEEHELETGIKSKEARKYIFNcL	Phosphate
T508	DKRIpTPLKTLNHQFVTMTHLL	Phosphate
T897 & S899	DQHQPPIIPDpTPpSPPVSVITIRSdT	Phosphate

Table 3.2 Summary of HIPK1 PTMs identified from CAD and ETD analysis of the initial Asp-N digestion.

3.5.5 CAD and ETD Analysis of Asp-N/Trypsin HIPK1 Peptides

Due to the limited HIPK1 sequence coverage using endoproteinase Asp-N alone, multiple *in silico* sub-digestions were compared to increase coverage of peptides not previously detected. Sub-digestion with trypsin cleaved the largest peptides to lengths amenable for analysis. HPLC-MS/MS methods identical to those used previously were subsequently applied to the Asp-N/trypsin digest and once again, CAD and ETD analyses were used to generate complementary data sets for sequence and PTM identification. Following extensive data analysis, the overall sequence coverage increased to nearly 87% (**Figure 3.12**). Coverage of C-terminal and central HIPK1 regions increased dramatically.

```

1 MASQLQVFSP PSVSSSAFCS AKKLEIEPSG WDVSGOSSND KYYTHSKTLP
51 ATOGOANSSH QVANFNIPAY DOGLLLPAPA VEHIVVTAAD SSGSAATSTF
101 OSSOTLTHRS NVSLLEPYOK CGLKRKSEEV DSNGSVQIIE EHPPLMLQNR
151 TVVGAAATTT TVTTKSSSSS GEGDYQLVOH EILCSMTNSY EVLEFLGRGT
201 FGQVAKCWR STKEIVAIKI LKNHPSYARQ GQIEVSILSR LSSENLAGEYN
251 FVRSYECFQH KNHTCLVFEM LEONLYDFLK ONKFSPLPLK YIRPILQOVA
301 TALMKLKSLG LIHADLKPEN IMLVDPVRQP YRVKVIDFGS ASHVSKAVCS
351 TYLQSRYYRA PEIILGLPFC EAIDMWSLGC VIAELFLGWP LYPGASEYDQ
401 IRYISOTOGL PAEYLLSAGT KTTRFFNRDP NLGYPLWRLK TPEEHELETG
451 IKSKEARKYI FNCLDDMAQV NMSTDLEGTD MLAEKADRRE YIDLLKMLT
501 IDADKRITPL KTLNHQFVTM THLLDFPHSN HVKSCFONME ICKRRVHMYD
551 TVSQIKSPFT THVAPNTSTN LTMSFSNQLN TVHNQASVLA SSSTAAAATL
601 SLANSDVSL NYQSALYPSS AAPVPGVAQQ GVSLOPGTTO ICTQTDPFQQ
651 TFIVCPPAFO TGLOATTKHS GFPVRMDNAV PIVPOAPAAO PLQIQSGVLT
701 QGSCTPLMVA TLHPQVATIT PQYAVPFTLS CAAGR PALVE QTA AVLQAWP
751 GGTQOILLPS TWOQLPGVAL HNSVQPTAMI PEAMGSGOOL ADWRNAHSHG
801 NOYSTIMOQP SLLTNHVTLA TAQPLNVGVA HVVROOQSSS LPSKKNKQSA
851 PVSSKSSLDV LPSQVYSLVG SSPLRTTSSY NSLVPVQDQH QPIIIPDTPS
901 PPVSVITIRS DTDEEEDNKY KPSSSGLKPR SNVISYVTVN DSPDSDSSLS
951 SPYSTDTLSA LRGNSGSVLE GPGRVVADGT GTRTIIVPPL KTQLGDCTVA
1001 TOASGLLSNK TKPVASVSGQ SSGCCITPTG YRAORGGTSA AQPLNLSQNO
1051 OSSAAPT SQE RSSNPAPRRQ QAFVAPLSQA PYTFQHG SPL HSTGHPHLAP
1101 APAHLP SQAH LYTYAAPTSA AALGSTSSIA HLFSPQGS SR HAAAYTTHPS
1151 TLVHQVPVSV GPSLLTSASV APAQYQHQFA TQSYIGSSRG STIYTGYP LS
1201 PTKISQYSYL

```

Figure 3.12 Proteolysis using Asp-N and trypsin sub-digestion results in nearly 87% HIPK1 sequence coverage. Residues that were unambiguously identified are underlined.

Upon inspection of the data, a second peptide was identified to possibly contain *O*-GlcNAc. The accurate mass of the precursor ion corresponded to singly GlcNAcylated and carbamidomethylated peptide DcTVATQASGLLSNK within 1.8 ppm. The elution profile and CAD spectrum also exhibited the expected patterns. A data dependent ETD MS/MS of this species confirmed GlcNAcylation on T1001 (**Figure 3.13**). The modification was estimated to exist at 6.5% relative abundance.

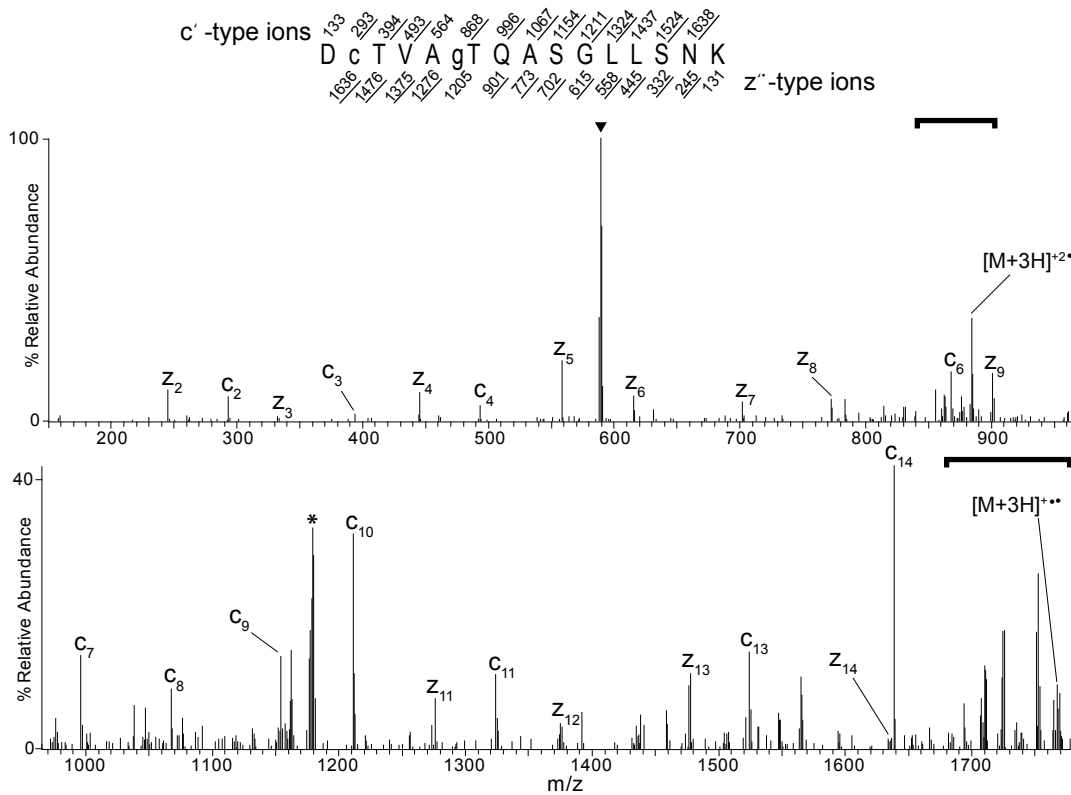


Figure 3.13 ETD MS/MS spectrum of singly GlcNAcylated, carbamidomethylated HIPK1 peptide, DcTVATQASGLLSNK. Fragment ions detected are labeled within the spectrum and underlined in the peptide sequence. Heavy bracketed ions correspond to neutral losses from labeled reduced charge species (RCS). RCS from species that coelute and fall within the precursor isolation window are labeled with *. The inverted triangle (▼) designates ions that fall within the precursor isolation window. Nearly all fragment ions expected are detected and demonstrate *O*-GlcNAcylation of T1001.

Further inspection of the data revealed additional HIPK1 phosphorylation sites. In addition to the five sites already identified, nine more were found on

peptides uniquely generated by trypsin sub-digestion (**Table 3.3**). Site

identifications were made possible by ETD and CAD MS/MS spectra. The

phosphorylation sites were present in largely varying amounts from 7 to 85%

relative abundance.

Site	Peptide Sequence	Modification	Relative Abundance
S9	DDDDKLAAMASQLQVFpSPPSVSSSAFCSAK	Phosphate	85%
T87	DQGLLLPAPAVEHIVVgTAA	<i>O</i> -GlcNAc	31.80%
S127	EPYQKCGLKRKpSEEV	Phosphate	31%
T441	DPNLGYPLWRLKpTPEEHELETGIKSKEARKYIFNCL	Phosphate	24%
T508	DKRIpTPLKTLNHQFVTMTHLL	Phosphate	11%
S798	NAHpSHGNQYSTIMQQPSLLTNHVTLATAQPLNVGV AHVVR	Phosphate	7.60%
S871	DVLPSQVYSLVGpSSPLR	Phosphate	ND
S872	DVLPSQVYSLVGSpSPLR	Phosphate	ND
T898	DQHQPIIIIPDpTPSPPVSVITIRSDT	Phosphate	ND
S900	DQHQPIIIIDTPpSPPVSVITIRSDT	Phosphate	ND
T912	SDpTDEEEDNKYKPSSSGLKPR	Phosphate	ND
S924	DEEEDNKYKPSpSSGLKPR	Phosphate	3%
S951	DSSLSpSPYSTDTLSALR	Phosphate	23%
S965	DTLSALRGNpSGSVLEGPGR	Phosphate	12%
T1001	DCTVAgTQASGLLSNK	<i>O</i> -GlcNAc	6.5%
S1200	GSTIYTGYPpSPTK	Phosphate	62%
S####	HAAAYTTHPSTLVHQPVP(SVGPSLLTS)*ASVAPAQY QHGFATQSYIGSSR	<i>O</i> -GlcNAc	ND*

Table 3.3 Summary of PTMs identified following HPLC-MS/MS analysis of Asp-N/trypsin digested HIPK1.

Peptides that coeluted in a single peak despite harboring two different post-translationally modified residues prevented the ability to determine relative abundance values (ND=not determined). A GlcNAc site near the C-terminus was unable to be accurately quantified due to chromatographical overlap of the modified and unmodified species, which resulted in ion suppression (ND*).

Evidence in the CAD analysis suggested a third possible GlcNAcylation site by accurate mass (-0.1 ppm) located near the C-terminus of HIPK1 on Asp-N/tryptic peptide, HAAAYTTHPSTLVHQVPVSVGPSLLTSASVAPAQYQHQFATQSYIGSSR. As expected, the CAD MS/MS spectrum was poor and provided little information to aid in the identification of the modification site. It did, however, contain GlcNAc-specific signature ions. ETD MS/MS analysis of the same species generated a spectrum that prevented unambiguous site identification. A few fragment ions suggested localization toward the center of the peptide (**Table 3.3**), but it was impossible to pinpoint the exact modification site. The poor quality of the acquired ETD spectra can be explained by two factors. First, the GlcNAcyated form of the large peptide was present at < 1 fmol. Peptides at very low abundance are likely to yield poorer MS/MS spectra due to significantly lower signal-to-noise ratios. Second, the m/z ratio of this tryptic peptide is not optimal for high quality ETD MS/MS spectra. Previous performance evaluations have determined that MS/MS provides maximum fragmentation when ETD is applied to multiply charged peptide cations where $z \geq +3$ and overall charge density is high (300 to 800 m/z) (45). ETD conditions are suboptimal for the Asp-N/tryptic peptide of interest because it exhibits m/z 892.783 in the most abundant charge state ($z=+6$) and contains no internal basic residues to encourage backbone cleavage.

There are two methods capable of circumventing these issues. First, a third proteolytic digestion can be used to shorten peptides to decrease the length and ideally, the precursor m/z . Second, charge state modification strategies can be

implemented to directly modify the m/z by increasing the overall charge state.

These methods rely on chemistries that add basic moieties to acidic side chains and C-termini. While these methods are very helpful for some sequences, the HIPK1 sequence in question includes no Asp and Glu residues. The only residue capable of being derivatized is the extreme C-terminal carboxylic acid. This would do very little to decrease the overall precursor m/z , so this option is not yet beneficial. Therefore, another sub-digest was used to modify the putative GlcNAcylated peptide. There are no internal acidic residues in the peptide to facilitate sub-digestion with Glu-C, so chymotrypsin was determined to be the optimum endoprotease because of several internal hydrophobic residues. Chymotrypsin is not commonly used in proteomic analyses because it cleaves at many residues with the most reliable cleavages occurring C-terminal to hydrophobic residues. In order to focus enzymatic cleavage to the most hydrophobic residues only, we limited the time of digestion to 5 h. Under these conditions, we expected to generate the following chymotryptic peptides: HAAAYTTHPSTL, VHQVPVSVGPSLL, TSASVAPAQY, QHQFATQSY, and IGSSR. Based on the ETD spectrum acquired on the intact peptide, we predicted that either VHQVPVSVGPSLL or TSASVAPAQY contained the GlcNAc moiety.

3.5.6 CAD Analysis of Asp-N/Chymotrypsin HIPK1 Peptides

The remaining ~60% of Flag-HIPK1 was sub-digested using chymotrypsin. An initial CAD MS analysis was performed on 1/6th the sub-digest, which

represented a similar aliquot to those previously analyzed. CAD was used to screen the sub-digest to ensure proper chymotryptic digestion and determine exactly where cleavages occurred in the peptide of interest. None of the 5 peptides expected following chymotryptic cleavage were expected to yield $z > +2$, so ETD analysis was not used. CAD analysis was optimal for this analysis due to the generation of GlcNAc-specific signature ions, which determined the GlcNAc-modified peptide. Following CAD analysis, we determined that chymotryptic cleavage occurred as expected. Manual analysis showed that VHQPVSVGPSLL was the only peptide that existed in an unmodified and singly GlcNAcylated form.

Figure 3.14A shows the characteristic elution profile, mass accuracy and charge state distribution for both forms. As expected, the singly GlcNAcylated form (m/z 767.9269) was exclusively in the +2 charge state, precluding interrogation by ETD.

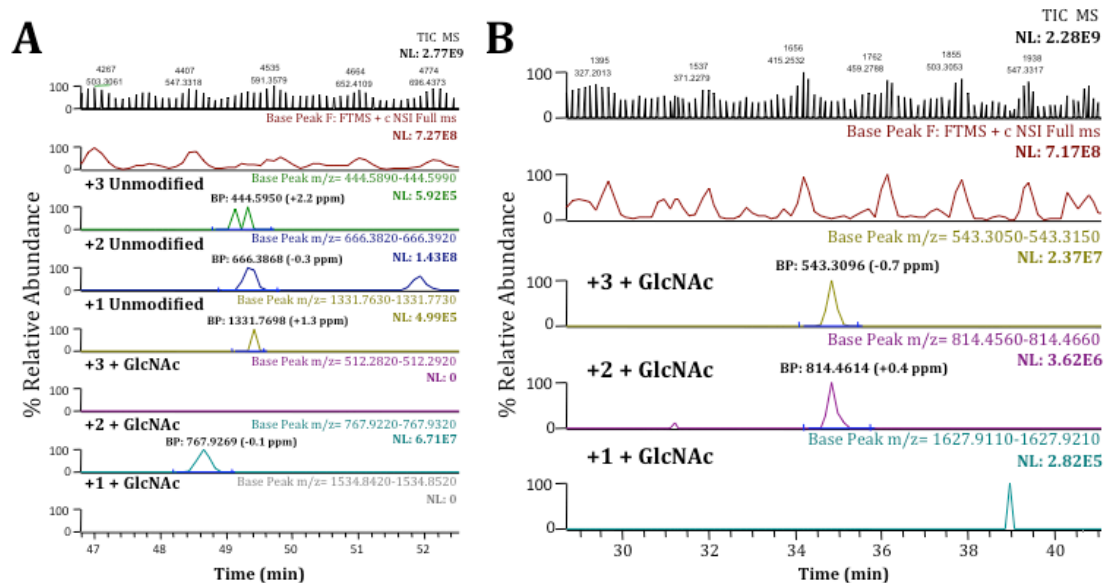


Figure 3.14 Charge state distribution of singly GlcNAcylated HIPK1 peptide VHQPVSVGPSLL. (A) Following chymotrypsin subdigestion, singly GlcNAcylated peptide VHQPVSVGPSLL precludes fragmentation by

ETD due to a +2 charge state. (B) After histamine derivatization of the C-terminus, the same peptide is 85% present in the +3 charge state, which provides conditions favorable for ETD.

CAD MS/MS of singly GlcNAcylated peptide VHQPVSVGPSLL showed two of the three expected signature ions (**Figure 3.15**). The low mass cutoff value prevented detection of the oxonium ion (m/z 204.1). Regardless, the presence of the other major ions sufficiently demonstrates the presence of GlcNAc. As usual, b- and y-type ions did not allow unambiguous identification of the GlcNAc site.

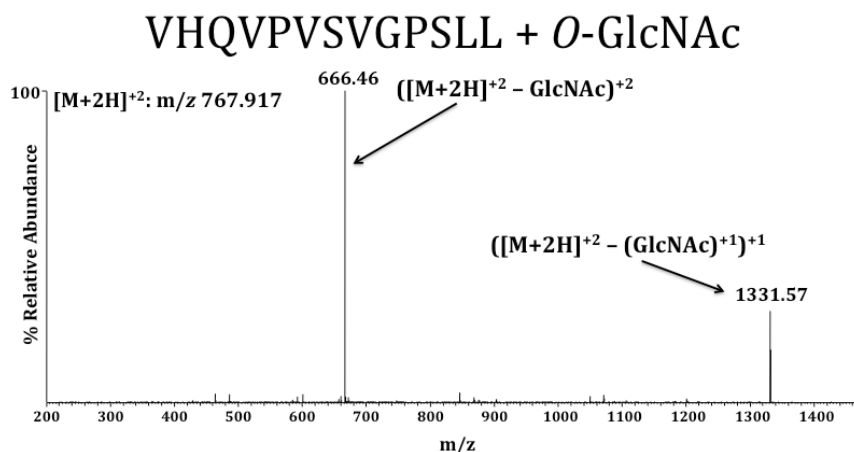


Figure 3.15 CAD MS/MS of GlcNAcylated HIPK1 peptide VHQPVSVGPSLL results in typical signature GlcNAc ion losses and minimal fragment ions.

3.5.7 Histamine Derivatization of HIPK1

In order to confidently determine the site of GlcNAcylation, we needed to increase the precursor charge state to at least +3 for successful interrogation with ETD. To achieve this, we used recently developed charge state modification strategies to replace C-terminal carboxyl groups and side chains of Glu and Asp with basic groups capable of protonation (46). These methods exploit the peptide coupling activity of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) to

activate carboxylic acid groups prior to attachment to primary amines. EDC is commonly used to facilitate attachment of primary amines from peptide N-termini or Lys side chains to carboxylic acid groups that cause intramolecular cyclization or peptide cross-linking (47, 48). Similarly, the methods implemented for HIPK1 use EDC to activate solvated carboxylic acid groups. However, a large excess (~1,000-fold) of histamine, which also includes a primary amine, is incubated with the activated peptides. Excess histamine encourages preferential amide linkage to carboxyl groups in place of amines from internal residues or those on other peptides (**Figure 3.16**). As a result, two products are formed: histamine-derivatized peptides and a urea-like byproduct. Derivatized peptides can undergo protonation at histamine groups to promote precursor charge state enhancement. Neither histamine nor the urea-like byproduct are retained by C18 due to small size and hydrophilicity. This allows loading of the reaction mixture directly onto the RP HPLC column without worry of contamination by reagents or byproducts.

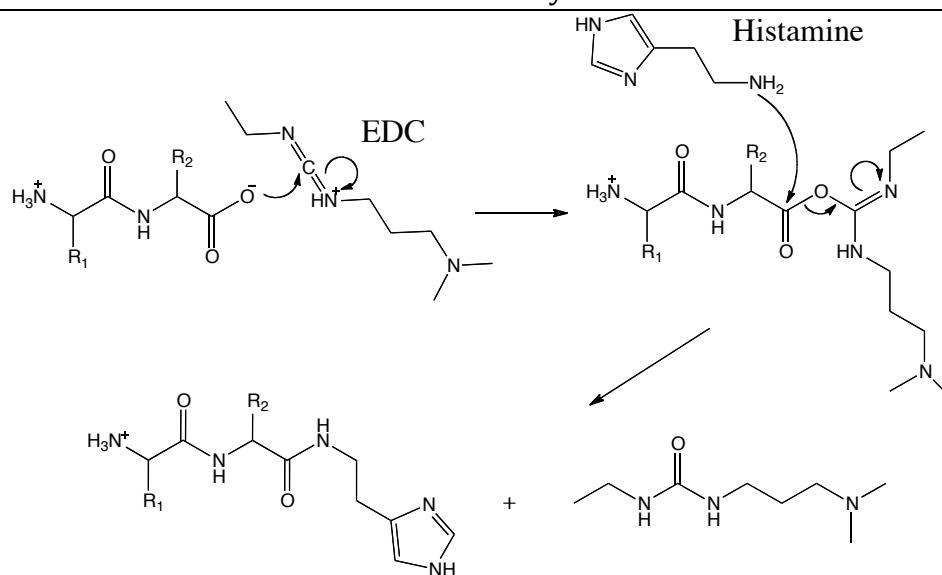


Figure 3.16 EDC-mediated peptide derivatization with histamine. Activation of the carboxylic acid group is mediated by EDC and allows histamine addition at the original carboxyl group to result in a derivatized peptide and a byproduct that is not retained by C18 resin. Single protonation of the histamine under RP-HPLC conditions can increase the overall precursor charge state by +1 per modified residue.

Figure 3.14B demonstrates the success of histamine-mediated charge state enhancement. GlcNAcylated VHQPVS VSGPSLL originally existed exclusively in the +2 charge state. Following derivatization with histamine, 85% of the precursor was in the +3 charge state. This allowed successful interrogation by ETD and unambiguous GlcNAc site assignment to S1159 (**Figure 3.17**).

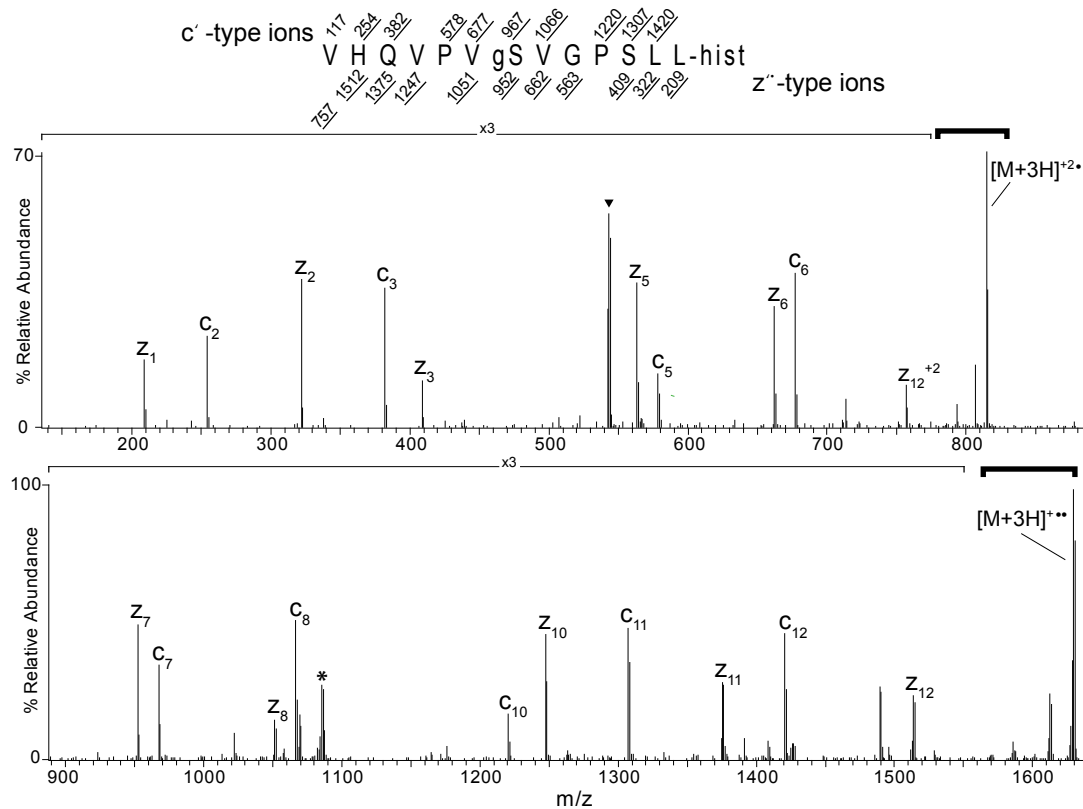


Figure 3.17 ETD MS/MS spectrum of singly GlcNAcylated and histamine-derivatized HIPK1 peptide, VHQPVS V G P S L L-hist. Histamine derivatization of this peptide alone allowed unambiguous site assignment to S1159. Fragment ions detected are labeled within the spectrum and underlined in the peptide sequence. Heavy bracketed ions correspond to neutral losses from labeled reduced charge species (RCS). RCS from species that coelute and fall within the precursor isolation window are labeled with *. The inverted triangle (▼) designates ions that fall within the precursor isolation window.

In addition to GlcNAcylated S1159, three additional phosphorylation sites were identified using Asp-N/chymotryptic digestion. Histamine derivatization allowed identification of phosphorylated S1088 exclusively. This site was found by increasing the charge state of peptide TFQHGpSPLHSTGHPH for subsequent analysis with ETD. As a result of chymotryptic sub-digestion, sequence coverage of HIPK1 consequently increased to greater than 97%. In total, 3 *O*-GlcNAc and 17 phosphorylation sites were identified on HIPK1 (**Table 3.4**). These 20 PTMs demonstrate that HIPK1 is highly modified. Prior to this work, only two

phosphorylation sites at Y352 and S872 were known, the latter of which we also identified in this study (35). These results demonstrate the utility of tailored MS analyses for PTM identification. As a result, we present a substantial contribution to the known PTM landscape of HIPK1.

Site	Peptide	Modification	Relative Abundance
S9	DDDDKLAAMASQLQVFpSPPSVSSSAFCSAK	Phosphate	85%
T87	DQGLLLPAPAVEHIVVgTAA	<i>O</i> -GlcNAc	31.8%
S127	EPYQKCGLKRKpSEEV	Phosphate	31%
T441	DPNLGYPLWRLKpTPEEHELETGIKSKEARKYIFNCL	Phosphate	24%
T508	DKRIpTPLKTLNHQFVTMTHLL	Phosphate	11%
S557	DTVSQIKpSPFTTH	Phosphate	< 9%
S798	NAHpSHGNQYSTIMQQPSLLTNHVTLATAQPLNVGVAHVVR	Phosphate	7.60%
S838-840	VVRQQQp(SSS)*LPSKKN	Phosphate	4.30%
S871	DVLPSQVYSLVGpSSPLR	Phosphate	ND
S872	DVLPSQVYSLVGSpSPLR	Phosphate	ND
T898	DQHQPPIIPDpTPSPPVSVITIRSDT	Phosphate	ND
S900	DQHQPPIIPDTPpSPPVSVITIRSDT	Phosphate	ND
T912	SDpTDEEEDNKYKPSGLKPR	Phosphate	ND
S924	DEEEDNKYKPSpSSGLKPR	Phosphate	3%
S951	DSSLSpSPYSTDTLSALR	Phosphate	23%
S965	DTLSALRGNpSGSVLEGPGR	Phosphate	12%
T1001	DCTVAgTQASGLLSNK	<i>O</i> -GlcNAc	6.50%
S1088	TFQHGpSPLHSTGHPH	Phosphate	<7%
S1159	VHQVPVgSVGPSLL	<i>O</i> -GlcNAc	<31%
S1200	GSTIYTGYPpSPTK	Phosphate	62%

Table 3.4 Summary of PTMs identified on HIPK1 using multiple proteases and histamine-derivatization. Peptides that coeluted in a single peak despite harboring two different post-translationally modified residues

prevented the ability to determine relative abundance values (ND=not determined). The abundance of pS1088 is estimated at less than 7% due to extensive chymotryptic cleavages of the unmodified species, which prevented detection of shorter unmodified forms that include S1088. The abundance of gS1159 is estimated at less than 31% due to coelution with the unmodified species and the resulting ion suppression.

3.5.8 Future Work

Currently, experiments are being conducted in the Hart laboratory to explore the role that *O*-GlcNAcylated HIPK1 plays in the Fas-mediated apoptotic pathway. Ron Copeland has determined that *O*-GlcNAcylation of WT HIPK1 decreases during Fas treatment of HEK293 cells. In order to determine which *O*-GlcNAc sites are relevant to this pathway, Ron has made T87A, T1001A and S1159A single site mutations. Interestingly, his pilot experiments demonstrate that overexpression of T87A HIPK1 alone causes increased pro-apoptotic caspase-3 activity as compared to HIPK WT. He has expressed the three single site HIPK1 mutations in HEK293 cells and initiated apoptosis by Fas treatment. Following immunoprecipitation of Flag-tagged HIPK1 mutants and Western blotting using the CTD 110.6 antibody, the T87A HIPK1 mutant was exclusively found maintain unchanged *O*-GlcNAcylation levels. Because HIPK1 WT exhibits overall decreases in *O*-GlcNAcylation following Fas treatment, this suggests that *O*-GlcNAcylation of T87 may be selectively removed during this pathway. Because of this, we hold a preliminary hypothesis that *O*-GlcNAcylation of T87 may play an inhibitory role in Fas-mediated apoptosis. This is further supported by increases in caspase-3 activity upon T87A mutation, which prevents *O*-GlcNAcylation of the site.

Prior to publishing this data, we hope to address several more questions including the effect that *O*-GlcNAcylation of S1159 has on HIPK1-ASK1 binding. *O*-GlcNAcylation of S1159, which is located in the ASK1-binding region, may influence HIPK1 binding to ASK1 to provide another opportunity for post-translational regulation of Fas-mediated apoptosis. We also want to investigate the interaction of HIPK1 with *O*-GlcNAcase during Fas-mediated apoptosis, which is suspected to occur based on observed decreases in overall HIPK1 GlcNAcylation. Additional studies remain to interrogate the direct effect of *O*-GlcNAcylation and/or phosphorylation on HIPK1 kinase function. Because *O*-GlcNAcylated T87 lies within the kinase domain, a definite possibility exists for regulation by OGT on HIPK1-mediated phosphorylation.

3.6 Conclusions

This work is the first attempt at comprehensive PTM analysis of human HIPK1. Characterization of 3 *O*-GlcNAc and 17 phosphorylation sites, 19 of which are novel, demonstrates not only that HIPK1 is highly modified, but also that the implemented analytical methods were highly efficient. Several factors contributed to the success of these methods. First, a highly pure HIPK1 sample obtained by anti-Flag purification provided several picomoles of protein for multiple tailored mass analyses. Second, the single and combinatorial proteolyses were advantageous for coverage of PTMs that were unidentified using a single protease. This is made evident by identification of multiple PTMs after each additional sub-digestion.

Third, CAD and ETD analyses provided complementary data sets that each contributed fragment ions for PTM site assignments at nearly every modified site. Lastly, charge state modification chemistry using histamine derivatization provided unambiguous localization of GlcNAcylated S1159 and phosphorylated S1088.

Aside from the analytical achievements of this study, we have also begun to investigate how GlcNAcylation affects HIPK1 function. Of the three novel GlcNAc sites identified, T87 is located within the kinase domain and T1001 and S1159 are located within the C-terminal domain. S1159 is located within the ASK1-binding region shown to facilitate apoptosis. Current experiments are aimed at defining the specific function of the three GlcNAc sites in particular in the hopes of elucidating how GlcNAcylation regulates Fas-mediated apoptosis. Preliminary experiments have demonstrated that overall HIPK1 WT GlcNAcylation levels significantly decrease during Fas-mediated apoptosis, which may suggest that GlcNAcylated HIPK1 may serve to inhibit the pathway. Because knowledge of HIPK1 function is severely limited compared to other kinases, our study stands to contribute substantially to the field.

3.7 References

1. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc, Torres, C. R. & Hart, G. W., Journal of Biological Chemistry **1984** (259): 3308-3317.
 2. Glycosylation of Nucleocytoplasmic Proteins: Signal Transduction and O-GlcNAc, Wells, L., Vosseller, K. & Hart, G. W., Science **2001** (291): 2376-2378.
 3. O-GlcNAc turns twenty: functional implications for post-translational modification of nuclear and cytosolic proteins with a sugar, Wells, L. & Hart, G. W., FEBS letters **2003** (546): 154-158.
 4. Cross-talk between GlcNAcylation and phosphorylation: roles in insulin resistance and glucose toxicity, Copeland, R. J., Bullen, J. W. & Hart, G. W., American Journal of Physiology - Endocrinology And Metabolism **2008** (295): E17-E28.
 5. Glycosylation of nuclear and cytoplasmic proteins. Purification and characterization of a uridine diphospho-N-acetylglucosamine:polypeptide beta-N-acetylglucosaminyltransferase. , Haltiwanger, R. S., Blomberg, M. A. & Hart, G. W., Journal of Biological Chemistry **1992** (267): 9005-9013.
 6. Purification and characterization of an O-GlcNAc selective N-acetyl-beta-D-glucosaminidase from rat spleen cytosol, Dong, D. L. & Hart, G. W., Journal of Biological Chemistry **1994** (269): 19321-19330.
 7. Akimoto, Y., Hart, G. W., Hirano, H. & Kawakami, H. (2005) in *Medical Molecular Morphology* Springer Japan, (38): 84-91.
 8. Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins, Hart, G. W., Housley, M. P. & Slawson, C., Nature **2007** (446): 1017-1022.
 9. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease, Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O., Annual Review of Biochemistry **2011** (80): null.
 10. Extensive Crosstalk Between O-GlcNAcylation and Phosphorylation Regulates Cytokinesis, Wang, Z., Udeshi, N. D., Slawson, C., Compton, P. D., Sakabe, K., Cheung, W. D., Shabanowitz, J., Hunt, D. F. & Hart, G. W., Sci. Signal. **2010** (3): ra2.
 11. The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny Shafi, R., Iyer, S. P. N., Ellies, L. G., O'Donnell, N., Marek, K. W., Chui, D., Hart, G. W. & Marth, J. D., Proceedings of the National Academy of Sciences of the United States of America **2000** (97): 5735-5739.
 12. The intersections between O-GlcNAcylation and phosphorylation: implications for multiple signaling pathways, Zeidan, Q. & Hart, G. W., J Cell Sci **2010** (123): 13-22.
-

13. O-Glycosylation of Nuclear and Cytosolic Proteins, Comer, F. I. & Hart, G. W., Journal of Biological Chemistry **2000** (275): 29179-29182.
 14. O-GlcNAc Transferase Is in a Functional Complex with Protein Phosphatase 1 Catalytic Subunits, Wells, L., Kreppel, L. K., Comer, F. I., Wadzinski, B. E. & Hart, G. W., Journal of Biological Chemistry **2004** (279): 38466-38470.
 15. Enrichment and Site Mapping of O-Linked N-Acetylglucosamine by a Combination of Chemical/Enzymatic Tagging, Photochemical Cleavage, and Electron Transfer Dissociation Mass Spectrometry, Wang, Z., Udeshi, N. D., O'Malley, M., Shabanowitz, J., Hunt, D. F. & Hart, G. W., Molecular & Cellular Proteomics **2010** (9): 153-160.
 16. Enrichment of O-GlcNAc Modified Proteins by the Periodate Oxidation-Hydrazide Resin Capture Approach, Klement, E., Lipinski, Z., Kupihar, Z., Udvardy, A. & Medzihradszky, K. F., Journal of Proteome Research **2010** (9): 2200-2206.
 17. Purification and Identification of O-GlcNAc-Modified Peptides Using Phosphate-Based Alkyne CLICK Chemistry in Combination with Titanium Dioxide Chromatography and Mass Spectrometry, Parker, B. L., Gupta, P., Cordwell, S. J., Larsen, M. R. & Palmisano, G., Journal of Proteome Research **2010**.
 18. Kim, Y. C., Udeshi, N., Balsbaugh, J., Shabanowitz, J., Hunt, D. & Olszewski, N., Amino Acids **2010** (40): 869-876.
 19. Mapping Sites of O-GlcNAc Modification Using Affinity Tags for Serine and Threonine Post-translational Modifications, Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J. & Hart, G. W., Molecular & Cellular Proteomics **2002** (1): 791-804.
 20. O-GlcNAc Regulates FoxO Activation in Response to Glucose, Housley, M. P., Rodgers, J. T., Udeshi, N. D., Kelly, T. J., Shabanowitz, J., Hunt, D. F., Puigserver, P. & Hart, G. W., Journal of Biological Chemistry **2008** (283): 16283-16292.
 21. A PGC-1 α -O-GlcNAc Transferase Complex Regulates FoxO Transcription Factor Activity in Response to Glucose, Housley, M. P., Udeshi, N. D., Rodgers, J. T., Shabanowitz, J., Puigserver, P., Hunt, D. F. & Hart, G. W., Journal of Biological Chemistry **2009** (284): 5148-5157.
 22. Homeobox genes and axial patterning, McGinnis, W. & Krumlauf, R., Cell **1992** (68): 283-302.
 23. Drosophila NK-homeobox genes Kim, Y. & Nirenberg, M., Proceedings of the National Academy of Sciences **1989** (86): 7716-7720.
 24. Congenital Heart Disease Caused by Mutations in the Transcription Factor NKX2-5, Schott, J.-J., Benson, D. W., Basson, C. T., Pease, W., Silberbach, G. M., Moak, J. P., Maron, B. J., Seidman, C. E. & Seidman, J. G., Science **1998** (281): 108-111.
 25. Homeodomain-interacting Protein Kinases, a Novel Family of Co-repressors for Homeodomain Transcription Factors, Kim, Y. H., Choi, C. Y., Lee, S.-J., Conti, M. A. & Kim, Y., Jour of Biological Chemistry **1998** (273): 25875-25879.
-

26. Characterization of Human Homeodomain-interacting Protein Kinase 4 (HIPK4) as a Unique Member of the HIPK Family, He, Q., Shi, J., Sun, H., An, J., Huang, Y. & Sheikh, M., Molecular and cellular pharmacology **2010** (2): 61-68.
 27. Roles of HIPK1 and HIPK2 in AML1- and p300-dependent transcription, hematopoiesis and blood vessel formation, Aikawa, Y., Nguyen, L. A., Isono, K., Takakura, N., Tagata, Y., Schmitz, M. L., Koseki, H. & Kitabayashi, I., **2006** (25): 3955-3965.
 28. Overlapping Roles for Homeodomain-Interacting Protein Kinases Hipk1 and Hipk2 in the Mediation of Cell Growth in Response to Morphogenetic and Genotoxic Signals, Isono, K., Nemoto, K., Li, Y., Takada, Y., Suzuki, R., Katsuki, M., Nakagawara, A. & Koseki, H., Mol. Cell. Biol. **2006** (26): 2758-2771.
 29. Homeodomain-Interacting Protein Kinase 1 Modulates Daxx Localization, Phosphorylation, and Transcriptional Activity, Ecsedy, J. A., Michaelson, J. S. & Leder, P., Mol. Cell. Biol. **2003** (23): 950-960.
 30. Modulation of the beta-Catenin Signaling Pathway by the Dishevelled-Associated Protein Hipk1, Louie, S. H., Yang, X. Y., Conrad, W. H., Muster, J., Angers, S., Moon, R. T. & Cheyette, B. N. R., PLoS ONE **2009** (4): e4310.
 31. HIPK1 interacts with c-Myb and modulates its activity through phosphorylation, Matre, V., Nordgard, O., Alm-Kristiansen, A. H., Ledsaak, M. & Gabrielsen, O. S., Biochemical and Biophysical Research Communications **2009** (388): 150-154.
 32. Tumor Necrosis Factor alpha-induced Desumoylation and Cytoplasmic Translocation of Homeodomain-interacting Protein Kinase 1 Are Critical for Apoptosis Signal-regulating Kinase 1-JNK/p38 Activation, Li, X., Zhang, R., Luo, D., Park, S.-J., Wang, Q., Kim, Y. & Min, W., Journal of Biological Chemistry **2005** (280): 15061-15070.
 33. DJ-1 interacts with HIPK1 and affects H₂O₂-induced cell death, Sekito, A., Koide-Yoshida, S., Niki, T., Taira, T., Iguchi-Arigo, S. M. M. & Arigo, H., Free Radical Research **2006** (40): 155-165.
 34. SENP1 mediates TNF-induced desumoylation and cytoplasmic translocation of HIPK1 to enhance ASK1-dependent apoptosis, Li, X., Luo, Y., Yu, L., Lin, Y., Luo, D., Zhang, H., He, Y., Kim, Y.-O., Kim, Y., Tang, S. & Min, W., **2008** (15): 739-750.
 35. Quantitative Phosphoproteomic Analysis of T Cell Receptor Signaling Reveals System-Wide Modulation of Protein-Protein Interactions, Mayya, V., Lundgren, D. H., Hwang, S.-I., Rezaul, K., Wu, L., Eng, J. K., Rodionov, V. & Han, D. K., Sci. Signal. **2009** (2): ra46.
 36. Kinase-Selective Enrichment Enables Quantitative Phosphoproteomics of the Kinome across the Cell Cycle, Daub, H., Olsen, J. V., Bairlein, M., Gnad, F., Oppermann, F. S., Körner, R., Greff, Z., Kéri, G., Stemmann, O. & Mann, M., Molecular Cell **2008** (31): 438-448.
-

37. Global Survey of Phosphotyrosine Signaling Identifies Oncogenic Kinases in Lung Cancer, Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., Hu, Y., Tan, Z., Stokes, M., Sullivan, L., Mitchell, J., Wetzell, R., MacNeill, J., Ren, J. M., Yuan, J., Bakalarski, C. E., Villen, J., Kornhauser, J. M., Smith, B., Li, D., Zhou, X., Gygi, S. P., Gu, T.-L., Polakiewicz, R. D., Rush, J. & Comb, M. J., *Cell* **2007** (131): 1190-1203.
 38. O-Linked beta-N-Acetylglucosaminyltransferase Substrate Specificity Is Regulated by Myosin Phosphatase Targeting and Other Interacting Proteins, Cheung, W. D., Sakabe, K., Housley, M. P., Dias, W. B. & Hart, G. W., *Journal of Biological Chemistry* **2008** (283): 33935-33941.
 39. Roles of the Tetratricopeptide Repeat Domain in O-GlcNAc Transferase Targeting and Protein Substrate Specificity, Iyer, S. P. N. & Hart, G. W., *Journal of Biological Chemistry* **2003** (278): 24608-24616.
 40. Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry, Udeshi, N. D., Compton, P. D., Shabanowitz, J., Hunt, D. F. & Rose, K. L., *Nature Protocols* **2008** (3): 1709-1717.
 41. Structure-based Design of beta-1,4-Galactosyltransferase I (Gal-T1) with Equally Efficient N-Acetylgalactosaminyltransferase Activity, Ramakrishnan, B. & Qasba, P. K., *Journal of Biological Chemistry* **2002** (277): 20833-20839.
 42. Determination and removal of impurities in 2-D LC-MS of peptides, Mihailova, A., Lundanes, E. & Greibrokk, T., *Journal of Separation Science* **2006** (29): 576-581.
 43. Simple and universal tool to remove on-line impurities in mono- or two-dimensional liquid chromatography-mass spectrometry analysis, Hesse, A.-M., Marcelo, P., Rossier, J. & Vinh, J., *Journal of Chromatography A* **2008** (1189): 175-182.
 44. N-glycosylation and the production of recombinant glycoproteins, Parekh, R. B., Dwek, R. A., Edge, C. J. & Rademacher, T. W., *Trends in Biotechnology* **1989** (7): 117-122.
 45. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry, Good, D. M., Wirtala, M., McAlister, G. C. & Coon, J. J., *Molecular & Cellular Proteomics* **2007** (6): 1942-1951.
 46. Exploration and Enhancement of Enzymatic and Chemical Peptide Modification Strategies for Optimization Fragmentation by Electron Transfer Dissociation, English, A. M., Balsbaugh, J. L., Shabanowitz, J. & Hunt, D. F., **2010**, 58th ASMS Conference on Mass Spectrometry and Allied Topics, Salt Lake City, UT, American Society for Mass Spectrometry.
 47. Techniques in Protein Modification, Lundblad, R. **1995**, CRC Press, 24-244.
 48. Recent development of peptide coupling reagents in organic synthesis, Han, S. Y. & Kim, Y. A., *Tetrahedron* **2004** (60): 2447-2467.
-

4 Discovery of Novel Protein Targets of α -N-terminal Methylation and N-terminal RCC1 Methyltransferase

4.1 Abstract

Post-translational methylation has been documented on numerous proteins and perhaps most extensively on highly modified histone proteins. With few exceptions, the majority of reported eukaryotic protein methylation has been found on Arg and Lys side chains. However, α -methylation of protein N-termini has also been identified in eukaryotes. Since its initial discovery more than 30 years ago, virtually no new substrates of N-terminal methylation have been documented. Recently, regulator of chromatin condensation 1 (RCC1) has been identified to be α -N-methylated, and additionally, functional significance has been attributed to the rare PTM for the first time. The generation of an anti-me²-SPK antibody allowed screening of HeLa extracts to identify fractions that contained α -N-methyltransferase activity. Active fractions were found to contain the methyltransferase, METTL11a. Evaluation of N-terminal methyltransferase activity towards RCC1 resulted in the discovery that METTL11a, later named N-terminal RCC1 methyltransferase (NRMT), is responsible for catalyzing α -N-methylation. For the first time, enzymology behind α -methylation was known. Retinoblastoma protein and SET were subsequently identified as NRMT substrates but information regarding proteome-wide α -methylation remained to be shown. In this chapter, we present novel methods for the identification of N-terminal methylated proteins involving immunopurification and protein identification using HPLC in tandem with

high resolution mass spectrometry implementing collision-activated and electron transfer dissociation.

4.2 Introduction

4.2.1 Protein Methylation by Methyltransferases

The covalent attachment of a methyl group from S-adenosyl methionine (SAM) to proteins represents one of the most common PTM events in biological systems (1). Methylation of nitrogen on amino and guanidino groups of Lys and Arg side chains, respectively, has been extensively reported. The most heavily studied methylated proteins are histones that contain multiple Arg and Lys residues in highly modified N-terminal regions. Protein methylation is catalyzed by a group of enzymes called methyltransferases (MTs). It has been estimated that 1% of all genes within the mammalian genome encode for MTs. The majority of mammalian MTs modify DNA and small molecules whereas only a fraction modify Lys and Arg side chains (2). These residues can be completely unmodified, monomethylated (me) or dimethylated (me₂). Lys residues can also be trimethylated (me₃) (1, 3). The diversity that results from variable methylation is further increased by protein arginine MTs (PRMTs) that perform either symmetric or asymmetric guanidino methylation (2). Methylation can dramatically change protein character by increasing hydrophobicity and steric bulk (4). This can consequently inhibit or enhance protein-protein interactions through specific recognition of the post-translationally modified residue. For example, chromodomains and Tudor domains

have been identified to recognize specific methylated lysines and arginine residues, respectively (3). In contrast to PTMs with shorter half lives like phosphorylation, methylation is usually longer lived despite the recent discovery of demethylases, which have been shown to remove methyl groups (1). Methylation has garnered great amounts of attention due to its ability to affect protein function and epigenetic regulation of gene expression (1-3). MS has been at the forefront of the identification and characterization of sites of protein methylation. Current high resolution mass spectrometric techniques provide mass accuracy that readily distinguishes between acetylation (+ 42.0105 Da) and trimethylation (+42.0468 Da). Despite vast efforts focused on characterizing protein methylation with MS, a unique type of methylation has not been as readily investigated since its discovery over 30 years ago.

4.2.2 α -N-Methylation of Protein N-termini

Post- and co-translational modification of protein N-termini have been known to exist for decades. N-terminal modification frequently involves cleavage of the initial Met residue. This is thought to correlate with the size of the second amino acid, specifically when residue 2 is one of 7 amino acids with the smallest radii of gyration (1, 5, 6). N-terminal acetylation is estimated to occur on nearly 85% of all eukaryotic proteins and is controlled by acetyltransferases that most commonly modify N-terminal Ala, Ser and Met residues in mammals (7). N-terminal modification can also involve the addition of myristoyl fatty acid groups for

membrane anchoring (5). Additionally, methylation of α -amino groups has been reported on only a few eukaryotic proteins since its discovery in 1976. *E. coli* ribosomal proteins L16 and L33 were first reported to be monomethylated on N-terminal Met residues (8, 9). Shortly thereafter, dimethylproline was discovered on the N-terminus of cytochrome c557 in *Crithidia oncopelti* (10). By 1987, only 10 more proteins were shown to exhibit N-terminal methylation. Rabbit myosin light chains (MLCs) 1 and 2 were the only mammalian proteins shown to be N-terminally methylated (11, 12). These studies suggested that only eukaryotic N-terminal sequences APK- and PPK- were modifiable although sequences beginning with Met, Ala and Phe were found to be modified as well in lower organisms (11). Since the 1980s, there had not been a broad N-methylation discovery. However, the field dramatically progressed recently due to identification of N-terminal methylation on human regulator of chromatin condensation 1 (RCC1).

4.2.3 RCC1 α -N-Methylation

RCC1 is the nucleotide exchange factor for the Ran GTPase, is required for Ran nuclear translocation and binds chromatin to promote mitotic spindle organization and nuclear envelope formation (13-16). Essential to spindle organization is proper RCC1 binding to chromatin, a structure composed of DNA and histone proteins that facilitates DNA compaction (13-15). The N-terminal RCC1 tail, which resembles histone tails in basicity and PTM abundance, is required for DNA binding. A separate region of RCC1 binds histones H2A and H2B (13, 17).

Following Met cleavage, all mammalian RCC1 N-termini begin with Pro except for primate RCC1, which begins with Ser (14). Further interrogation of PTMs on human RCC1 identified α -mono-, di- and trimethylation on the N-terminal Ser (14). Site mutations of the N-terminal residue suggested that three sequences allow efficient N-terminal methylation, (Ser/Pro/Ala)-Pro-Lys, which agrees with previously known mammalian α -methylation targets (11, 12, 14). This also suggested that perhaps all mammalian RCC1 (Ser or Pro) N-termini are capable of being methylated. Chen et al. determined that disruption of N-terminal methylation by sequence mutations resulted in mislocalization of RCC1 from chromatin during mitosis, which then led to disorganized spindle assembly (**Figure 4.1**) (14). It is believed that mislocalization of RCC1 occurs due to the disruption of the association of the N-terminal tail of RCC1 with chromatin when N-terminal methylation is absent.

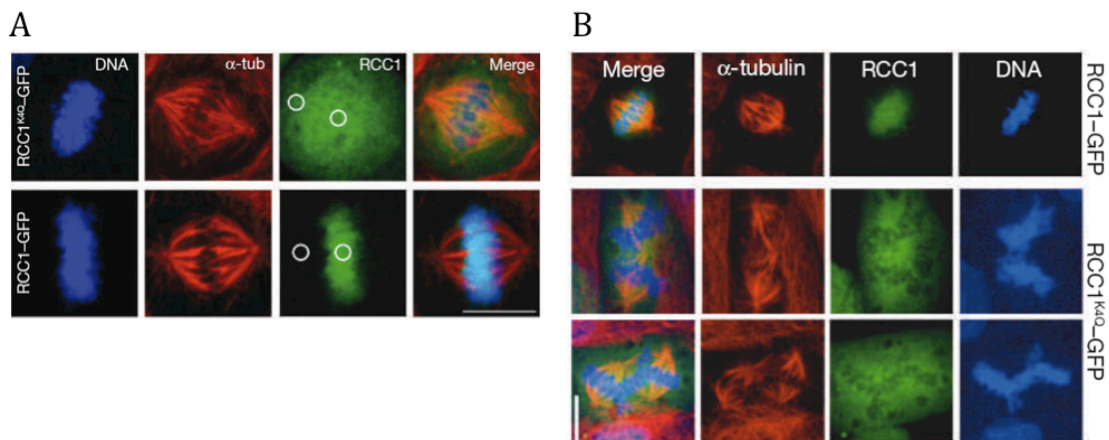


Figure 4.1 N-terminal methylation of GFP-tagged RCC1 regulates interactions with chromosomes and mediates spindle pole organization. (A) Mutations in the N-terminal sequence of RCC1 inhibit methylation (K4Q) and result in the mislocalization of RCC1 in mitotically arrested MDCK cells. (B) Spindle pole disorganization results from the same K4Q RCC1 mutations that promote loss of N-terminal methylation. Figure reproduced with permission from (14).

Methylation-enhanced binding by RCC1 is thought to rely on the fixed charge

associated with fully methylated N-termini (me3-APK, me3-SPK and me2-PPK)

(Figure 4.2). The fixed charge is expected to electrostatically favor negatively charged DNA to enhance chromatin binding. N-terminal tail localization to DNA and RCC1 body attachment to histones H2A and H2B are supported by structural data (14, 15). These results ascribe functionality of α -N-terminal methylation for the first time.

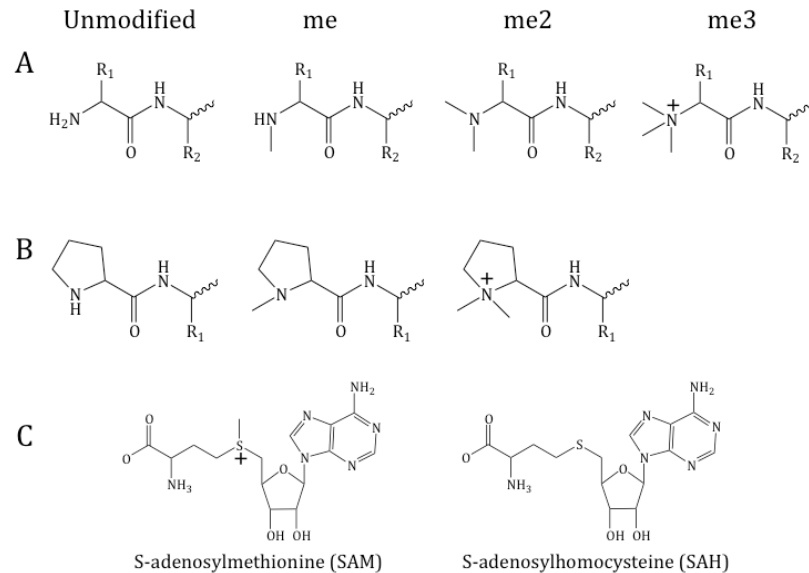


Figure 4.2 Full methylation of protein N-termini leads to a fixed positive charge on the α -amino group. (A) Trimethylation of any non-Pro N-terminus results in a fixed charge, while (B) N-terminal Pro residues can only facilitate dimethylation. (C) SAM acts as a methyl donor via thiol reduction.

4.2.4 N-terminal RCC1 Methyltransferase (NRMT) and Its Substrates

The discovery of N-terminal methylation on RCC1 prompted further study into the enzymology behind the modification. Previously, classification of hypothetical MTs was based exclusively on sequences of known substrates (11).

Recently, antibodies were generated to N-terminally modified RCC1 sequences me3-

Ser and me²-Pro (18). Fractionated HeLa lysates were assayed for N-terminal methylation using these antibodies to signal active fractions for MS analysis. Mass spectral data identified methyltransferase METTL11a in active fractions (18). METTL11a, now named NRMT, was shown to modify SPK and PPK sequences *in vitro* using Flag-purified NRMT and His-tagged RCC1 and *in vivo* using overexpressed NRMT in HEK 293LT cells (18). Lentiviral depletion of NRMT resulted in significantly decreased methylation of RCC1 without influencing overall RCC1 expression levels (18). The crystal structure of NRMT was solved with bound SAM and provided a structure for substrate modeling using ICM-PRO. An RCC1-like peptide was used for substrate modeling (XPKRIA, where X=Ala, Pro, Ser). From this data, NRMT was found to contain a large chromodomain-like cavity that facilitates binding of modeled peptides (18) (**Figure 4.3A**). Although all possible peptides were not modeled, mutation of the RCC1 N-terminal residue to all 20 amino acids demonstrated that NRMT-mediated methylation varied accordingly (**Figure 4.3B**). While acidic and hydrophobic residues yielded decreased methylation, sequences with Ala/Ser/Pro-Pro-Lys consensus sequences were shown to be optimal.

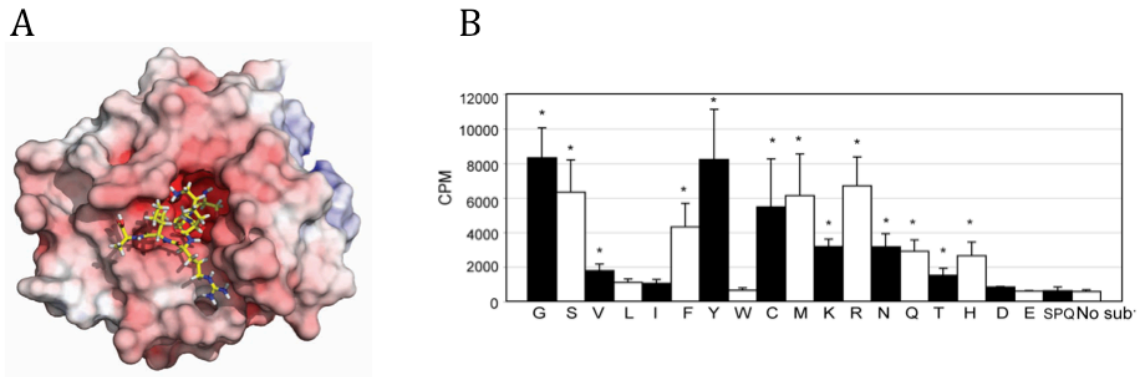


Figure 4.3 Interrogation of NRMT structure and specificity. (A) Crystal structure of NRMT with plotted electrostatic potential demonstrates that binding of SPKRIA RCC1-like peptide takes place within a large hydrophobic cavity similar to chromodomains. (B) Mutation of the N-terminal residue of His-RCC1 results in varied levels of methylation incorporation measured using the incorporation of ^3H -SAM. Figures reproduced with permission from (18).

Because N-terminal methylation has a specific function on RCC1, it stands to reason that other proteins may also require N-terminal methylation for binding purposes. An initial attempt to purify NRMT substrates using anti-me3-SPK IP from HeLa lysates provided identification of only one protein, patient SE translocation (SET). IP and mass spectrometric analysis of SET α , the only splice variant with an (Ala-Pro-Ser)-Pro-Lys N-terminal sequence, confirmed that nearly 100% of SET α was α -N-trimethylated (18). After the discovery of only one new NRMT substrate using this IP, a list was compiled of possible substrates based on the NRMT consensus sequence. Retinoblastoma protein (RB) was included in this list. Recombinant NRMT was later found to *in vitro* methylate RB while endogenous RB in the colon cancer cell line HCT116 was found to contain N-terminal dimethylation on Pro using immunofluorescence (18).

RCC1, SET and RB constitute the first three known human NRMT substrates.

While these merely begin to expand knowledge of NRMT activity in the proteome, it is clear that new methodologies are needed to progress these efforts further. A GenBank list of proteins that contain NRMT consensus sequences suggests more targets are possible (**Table 4.1**). This chapter presents new methods for elucidating novel NRMT substrates using IP and high resolution MS and CAD and ETD MS/MS. As a result, several new α -N-methylated mouse proteins were identified.

Accession	Gene Name	Accession	Gene Name
	SPK		APK
CAA46340	zinc-finger protein (ZNFpT17)	NP_000423	slow cardiac myosin regulatory light chain 2
NP_057463	VCX-variable charge testis protein	NP_524144	fast skeletal myosin alkali light chain 1
NP_001260	RCC1- regulator of chromosome condensation	CAG33244	fast skeletal myosin light chain 2
AAL86399	transmembrane channel-like protein 1	CAA36256	myosin, light chain 4, alkali; atrial, embryonic
	PPK	NP_006800	translocase of outer mitochondrial membrane 34
BAD97233	X-prolyl aminopeptidase 1, soluble variant	NP_783313	NADH-ubiquinone oxidoreductase subunit
NP_003753	vesicle-associated membrane protein 4	NP_009038	translocon-associated protein gamma subunit
AAO73048	vaccinia related kinase 2	BAD12560	damage-specific DNA binding protein 2
AAN64133	retinoblastoma 1	NP_996879	sperm associated antigen 17
NP_060941	erythropoietin 4 immediate early response	AAW51946	chemokine-like factor super family 2
NP_001019	ribosomal protein S25	NP_061149	TCP11
NP_000967	ribosomal protein L12	NP_000975	ribosomal protein L23a
AAI27120	XIRP1 protein	AAI41841	mortality factor 4 like 1
NP_037473	Obg-like ATPase 1	NP_037490	peptidylarginine deiminase type I
NP_002466	smooth muscle and non-muscle myosin alkali light chain 6B	NP_001472	growth arrest-specific 8
	MPK	AAD29855	NAD ⁺ ADP-ribosyltransferase 3
CAB90453	human non histone chromosomal protein HMG 14	NP_150092	gamma-aminobutyric acid (GABA) A receptor
AAP35727	eukaryotic translation initiation	AAL06239	testis development protein

	factor 1A		NYD-SP29
CAI20350	small nuclear ribonuclear polypeptide C	NP_001116293	SET
CAI22233	60S acidic ribosomal protein PO	NP_078782	galactosidase, beta 1-like
CAA43925	ribosomal protein L7a		
AAH01282	High mobility group nucleosomal binding domain 4		

Table 4.1 Probable NRMT substrates based on SPK, APK, PPK and PK N-terminal sequences following initiating Met cleavage. MPK sequences were considered based on previously known modified sequences. Accession numbers and sequences derived from GenBank and PubMed entries (NCBI).

4.3 Materials

4.3.1 Reagents

American Peptide (Sunnyvale, CA)

N-terminally dimethylated synthetic peptide, me₂-PPKRIAKRRSC

Anaspec, Inc. (San Jose, CA)

N-terminally trimethylated synthetic peptide, me₃-SPKRIAKRRSC

Mallinckrodt, Inc. (Paris, KY)

Acetonitrile, HPLC grade, $\geq 99.8\%$ purity

PQ Corporation (Valley Forge, PA)

Kasil® 1624

Lithisil®

Roche (Penzberg, Germany)

Endoproteinase Asp-N, sequencing grade

Endoproteinase Glu-C, sequencing grade

Santa Cruz Biotechnology, Inc. (Santa Cruz, CA)

Rabbit IgG

Sigma Aldrich (St. Louis, MO)

Ammonium bicarbonate

Angiotensin I acetate salt hydrate, $\geq 90\%$ purity (human)

Azulene

Dithiothreitol

Formamide

Glacial acetic acid, $\geq 99.9\%$ purity

Iodoacetamide

Tetramethylammonium silicate

Vasoactive intestinal peptide fragment 1-12, $\geq 97\%$ purity (human, porcine, rat)

Thermo Scientific (Rockford, IL)

Pierce® Crosslink Immunoprecipitation Kit (includes buffers, sample tubes and spin columns)

4.3.2 Equipment & Instrumentation

Agilent Technologies (Palo Alto, CA)

HP1100 high performance liquid chromatograph

Labconco Corporation (Kansas City, MO)

Centrivap centrifugal concentrator

Polymicro Technologies, LLC (Pheonix, AZ)

360 μm o.d. x 50 μm i.d. polyimide coated fused silica capillary

360 μm o.d. x 75 μm i.d. polyimide coated fused silica capillary

Sutter Instrument Co. (Navato, CA)

P-2000 microcapillary laser puller

Thermo Scientific (San Jose, CA/Bremen, Germany)

LTQ-Orbitrap hybrid mass spectrometer

LTQ-FT-ICR hybrid mass spectrometer

YMC Co., Ltd. (Kyoto, Japan)

C18 resin (5-20 μm diameter, 300 \AA pore size)

C18 resin (5 μm diameter, 120 \AA pore size)

0.060 in o.d. x 0.012 in i.d. Teflon tubing

4.4 Methods

The sample preparation and cell biology-based experiments were completed as described in (18) by Dr. Christine Schaner Tooley and Janusz Petkowski in Dr. Ian G. Macara's laboratory in the Department of Microbiology at the UVA School of Medicine.

4.4.1 Immunoblotting for me²-PPK & me³-SPK in *Mus musculus* Tissue

Approximately 20 μ g of protein from each of several mouse tissue lysates were subject to SDS-PAGE and transferred to nitrocellulose for Western blot analysis. Immunoblots were completed using anti-me²-PPK and anti-me³-SPK antibodies generated as previously described (18).

4.4.2 Anti-me³-SPK *Mus musculus* Immunoprecipitation

For isolation of N-terminally methylated substrates in *Mus musculus* heart and spleen tissues, 10 μ g of rabbit IgG and 10 μ g of purified anti-me³-SPK serum were covalently cross-linked to approximately 20 μ L Protein A/G on agarose using the Pierce® Crosslink IP Kit according to manufacturer's instruction and in (14).

For SDS-PAGE analysis, antigens were eluted from the antibody using 10 μ L Elution Buffer (pH 2.8) into a new tube using spin columns and centrifugation. Proteins were suspended in 1X Lane Marker Sample Buffer (0.3 M Tris-HCl, 5% SDS,

50% glycerol, lane marker tracking dye (pH 6.8), boiled, separated on polyacrylamide and detected with Coomassie staining. An identical anti-me³-SPK IP was prepared for MS analysis. After antigen retention on agarose, samples were washed with 1X TBS detergent free buffer (0.025 M Tris, 0.15 M NaCl, pH 7.2) and 100 mM ammonium bicarbonate (pH 8.0).

4.4.3 Proteolytic Digestions and Sample Preparation for Mass Analysis

Anti-me³-SPK immunoprecipitated proteins were placed in a molecular weight cutoff (MWCO) Pierce® spin column for MS sample preparation. The agarose was washed three times with 50 μ L aliquots of 100 mM ammonium bicarbonate (pH 8) and spun down at 3,000 $\times g$ to remove excess solvent and components of lysis buffer and to equilibrate sample pH. Reduction of spleen and heart protein Cys residues was performed in 200 μ L of 100 mM ammonium bicarbonate and 10 μ L of 0.58 mM dithiothreitol for 1.5 h at room temperature. Carbamidomethylation was achieved by the addition of a 10 μ L aliquot of 1.15 mM iodoacetamide. Bead samples were split in half, placed in identical spin filters and suspended in 200 μ L of 100 mM ammonium bicarbonate. Six-hour digestions were similar to those described in (19). Briefly, 50% heart protein fractions were proteolytically digested on the spin filter with approximately 6.0 ng of either endoproteinase Asp-N or Glu-C. Likewise, 50% spleen protein fractions were proteolytically digested using 18 ng of either endoproteinase Asp-N or Glu-C. Following digestion, unmodified Asp-N- and Glu-C-generated peptides were isolated

using centrifugation and acidified using glacial acetic acid to pH 3.5. N-terminal peptides retained by the antibody were eluted into a separate tube using Pierce® elution buffer (pH 2.8). All peptide samples were dried down, resuspended in 40 μ L 0.1% acetic acid and stored at -35°C until mass analysis.

4.4.4 HPLC Capillary Column Assembly

An 8 cm precolumn was constructed using 360 μm o.d. x 75 μm i.d. fused silica capillary with a 2 mm Kasil® 1624 frit and packed with 5 cm of irregular C18 reverse phase resin (5-20 μm diameter, 300 Å pore size). A 9 cm analytical column was constructed using 360 μm o.d. x 50 μm i.d. fused silica capillary containing a 1 mm Lithisil® frit. The analytical column was packed with 7 cm of regular C18 resin (5 μm diameter, 120 Å pore size) and equipped with a laser-pulled, integrated electrospray emitter tip as in (20). Both columns were prepped using several picomoles of angiotensin to optimize chromatographic separation.

4.4.5 HPLC-MS/MS Analysis

Mass spectrometric analysis of spleen and heart endoproteinase Glu-C digests required 25% and 50% aliquots, respectively. N-terminal peptides were pooled for simultaneous analysis. For elution analyses, 50% of the combined Asp-N and Glu-C digest was used. In each case, peptides were pressure loaded onto the precolumn at a flow rate of 1 $\mu\text{L}/\text{min}$. Following sample loading and loading of 250 fmol each of angiotensin (angio) and vasoactive intestinal peptide (vaso), a 0.1 M

acetic acid desalting rinse was completed for 10 min at similar flow rates. The precolumn was butt-connected to the analytical column using a 1 cm Teflon sleeve (0.060 in o.d. x 0.012 in i.d). Peptides were gradient eluted at a split flow-generated rate of 60 nL/min into the mass spectrometer using an Agilent 1100 chromatograph and linear LC gradient as follows: 0-60% Solvent B in 60 min, 60-100% Solvent B in 5 min, 100% Solvent B hold for 5 min, 100% - 0% B in 5 min, 100% Solvent A for 20 min (Solvent A: 0.1 M acetic acid in water and Solvent B: 70% acetonitrile in 0.1 M acetic acid in water).

Mass analyses involved high resolution (60,000 resolving power at m/z 400) MS^1 survey scans in the Orbitrap or FT-ICR cell, and low resolution, data dependent MS/MS scans in the linear ion trap. Unmodified peptide Glu-C digests used for source protein identification were analyzed using eight consecutive, data dependent ETD MS/MS events. N-terminal peptides were analyzed using a toggle method consisting of sequential data dependent CAD and ETD analyses for each precursor ion. Data dependence parameters for MS/MS data acquisition included monoisotopic precursor selection "on," repeat count of 2, repeat duration of 30 s, exclusion list duration of 20 s, and precursor isolation window of 3 m/z . ETD parameters included 35 ms reaction time, charge state rejection "on" for $z=+1, +2$, FTMS AGC target of $2e5$, ITMSⁿ AGC target of $1e4$, and azulene ETD reagent target of $2e5$. CAD/ETD toggle methods implemented charge state rejection for $z = +1$ only.

All data were searched using OMSSA (version 2.1.1) using MS/MS peak lists generated by Bioworks Browser (version 3.3.1 SP1) against the entire human/rat/mouse RefSeq database (downloaded June 2009). Searches of CAD and ETD data shared the following parameters: Asp-N or Glu-C enzyme specificity, E-value cutoff of "1," missed cleavage count of 3, variable modifications carbamidomethylation of Cys, oxidation of Met, mono-, di- and trimethylation of protein N-termini, phosphorylation of Ser, Thr and Tyr, and ± 0.01 Da and ± 0.35 Da for precursor and product ion masses, respectively. OMSSA provided removal of reduced charge species in ETD data. These results were used to guide data analysis but all spectra were manually confirmed to ensure confident assignments.

4.5 Results & Discussion

4.5.1 Immunoblotting for α -N-methylation in *Mus musculus* Tissues

Various mouse tissues were immunoblotted against antibodies that recognize me2-PPK and me3-SPK to determine where the highest abundance of N-terminally methylated proteins exist. Equal protein amounts (20 μ g) from mouse liver, brain, heart, skeletal muscle, lung, kidney, testis, uterus, and spleen were used (**Figure 4.4**). The most intense signals were located in anti-me3-SPK blots against heart and skeletal muscle tissues. The presence of multiple strong bands suggests that the highest number of modified proteins exists in spleen tissue. Heart, skeletal muscle and spleen tissues were considered optimal for analysis. However, there was not sufficient skeletal muscle tissue available at the time of IP for mass analysis.

Fortunately, several bands found in skeletal muscle are represented in heart and spleen tissue.

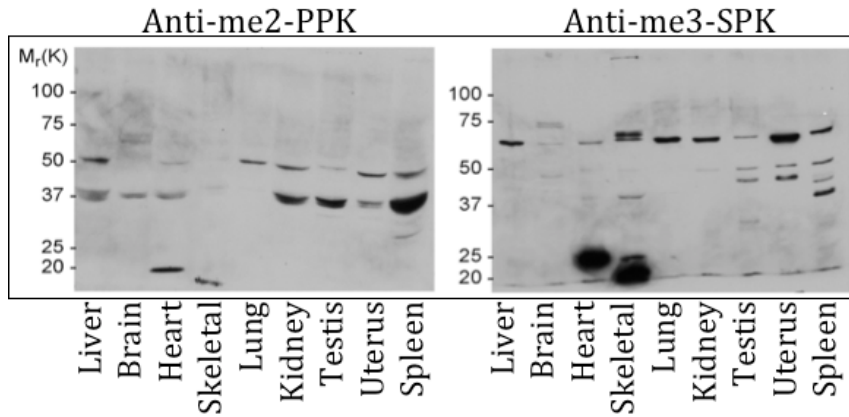


Figure 4.4 Immunoblotting against me2-PPK and me3-SPK in various *Mus musculus* tissue. Protein bands common to each lane suggest that some proteins with N-terminal methylation are ubiquitously expressed.

4.5.2 Anti-me3-SPK *Mus musculus* Heart & Spleen IP

Proteins were purified from mouse spleen and heart tissue lysates using an anti-me3-SPK antibody. Two IPs were simultaneously generated for SDS-PAGE (**Figure 4.5**) and MS analysis. SDS-PAGE displays several protein bands of varied molecular weight. In theory, all protein bands absent in the IgG control possess N-terminal methylation. The intensity of each Coomassie-stained tissue-specific band is estimated to represent approximate concentrations of 1 pmol or less.

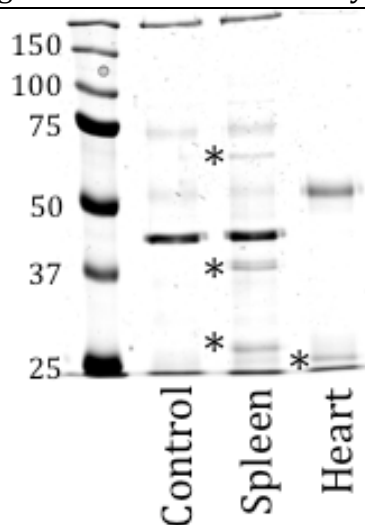


Figure 4.5 SDS-PAGE and Coomassie staining identifies proteins from spleen and heart tissue me³-SPK IPs. An IgG pulldown demonstrates that several proteins are specific to each tissue (denoted by “*”).

4.5.3 Novel LC-MS/MS Methods for Analysis of N-terminal Methylation

Due to the lack of preexisting methods for mass spectrometric characterization of N-terminally methylated proteins, we designed new methods for sample preparation and mass analysis. The choice of protease was made difficult for several reasons. First, the primary sequence for each protein was unknown and therefore, we could not choose an optimal endoproteinase to generate peptides amenable to HPLC-MS/MS analysis. Second, even though trypsin is commonly used for protein identification experiments, treatment of these potentially N-terminally methylated proteins with trypsin would produce short N-terminal peptides (SPK, PPK, APK, etc.). Peptides of this length are not typically retained under C18 reverse phase HPLC conditions. Identification of N-terminal methylation is only possible with MS/MS analysis of an efficiently retained, modified N-terminal peptide. The Genbank protein list was modified to include N-terminal amino acid sequences to

decide which endoprotease(s) best suits suspected targets (**Table 4.2**).

Endoproteases Glu-C and Asp-N were determined to be optimal because they mostly generate N-terminal peptides > 8-10 residues in length.

Accession	Gene	MW kD	N-terminal Sequence	Optimum Protease
SPK				
CAA46340	zinc-finger protein (ZNFpT17)	31.2	MSPKRDGLGTGDGLHSQVLQE.Q	Glu-C
NP_057463	VCX-variable charge testis protein	20.0	MSPKPRASGPPAKATE.A	Glu-C
NP_001260	RCC1- regulator of chromosome condensation	44.9	MSPKRIAKRRSPPA.D	Asp-N
AAL86399	transmembrane channel-like protein 1	87.7	MSPKKVQIKVEEKE.D	Asp-N/Glu-C
APK				
NP_000423	slow cardiac myosin regulatory light chain 2	18.7	MAPKKAKKRAGGANSNVFSMF E.Q	Glu-C
NP_524144	fast skeletal myosin alkali light chain 1	21.1	MAPKKDVKKPVAAAAAPAPA PAPAPAPAKPKEE.K	Glu-C
CAG33244	fast skeletal myosin light chain 2	19.0	MAPKRAKRRTVEGGSSVFSMF DQTQIQEFKE.A	Glu-C
CAA36256	myosin, light chain 4, alkali; atrial, embryonic	21.5	MAPKKPEPKKEAAKPMNVKML .D	Asp-N
NP_006800	translocase of outer mitochondrial membrane 34	34.5	MAPKFPDSVVEE.L	Glu-C
NP_783313	NADH-ubiquinone oxidoreductase subunit	14.8	MAPKVFRQYW.D	Asp-N
NP_009038	translocon-associated protein gamma subunit	21.0	MAPKGSSKQSEE.D	Asp-N
BAD12560	damage-specific DNA binding protein 2	26.7	MAPKKRPETQKTSE.I	Glu-C
NP_996879	sperm associated antigen 17	251.7	MAPKKEKGGTVNTSSKIWEPSL IAAQFNQN.D	Asp-N
AAW51946	chemokine-like factor super family 2	21.3	MAPKAAKGAKPEPAPPPPGA KPEE.D	Asp-N/Glu-C
NP_061149	TCP11	49.3	MAPKGILGSKPTAMNLSLE.G	Glu-C
NP_000975	ribosomal protein L23a	17.6	MAPKAKKEAPPKAE.A	Glu-C
AAI41841	mortality factor 4 like 1	37.2	MAPKQDPKPKFQE.G	Glu-C
NP_037490	peptidylarginine deiminase type I	74.6	MAPKRVVQLSLKMPHAVCVV GVE.A	Glu-C
NP_001472	growth arrest-specific 8	56.3	MAPKKKGKKGKAKGTPIV.D	Asp-N
AAD29855	NAD+ ADP-ribosyltransferase 3	60.1	MAPKPKPWVQTE.G	Glu-C
NP_150092	gamma-aminobutyric acid (GABA) A receptor	54.2	MAPKLLLLLCLFSGLHARSRKVE E.D	Asp-N/Glu-C
AAL06239	testis development protein NYD-SP29	102.8	MAPKQKKKTSRGKKRLKPVLAASE.D	Asp-N/Glu-C
NP_001116293	SET	33.4	MAPKRQSPLPPQKKKPRPPPAL GPPE.T	Glu-C
NP_078782	galactosidase, beta 1-like	74.1	MAPKKLSCLRLLLLPLSLTLLLP QA.D	Asp-N
PPK				
BAD97233	X-prolyl aminopeptidase 1, soluble variant	69.9	MPPKVTSELLRQLRQAMRNSE.Y	Glu-C

NP_003753	vesicle-associated membrane protein 4	16.3	MPPKFKRHLN.D	Asp-N
AAO73048	vaccinia related kinase 2	58.1	MPPKRNEKYKLPFPPEGKVL.D	Asp-N
AAN64133	retinoblastoma 1	106.1	MPPKTPRKTAATAAAAAAEP PPPPPPPEE.D	Asp-N/Glu-C
NP_060941	erythropoietin 4 immediate early response	48.6	MPPKKQAQAGGSKKAE.Q	Glu-C
NP_001019	ribosomal protein S25	13.7	MPPKDDKKKK.D	Asp-N
NP_000967	ribosomal protein L12	17.8	MPPKFKPNE.I	Glu-C
AAI27120	XIRP1 protein	56.3	MPPKKKQLPPKPAHLTQSHPP QRLPKPLPLSPSFSSE.V	Glu-C
NP_037473	Obg-like ATPase 1	44.7	MPPKKG.D	Asp-N
NP_002466	smooth muscle and non-muscle myosin alkali light chain 6B	22.7	MPPKDVVVKKPPAGPSISKPA KPAAAGAPPAKTA.E.P	Glu-C
MPK				
CAG46919	HMG2	9.3	MPKRKAEG.D	Asp-N
AAF03506	PDGFA associated protein 1	20.6	MPKGGRRKGGHKGRARQYTSPE E.I	Glu-C
CAB90453	human non histone chromosomal protein HMG 14	10.6	MPKRKVSSAE.G	Glu-C
AAP35727	eukaryotic translation initiation factor 1A	16.4	MPKNKGKGGKNRRRQKNE.N	Glu-C
CAI20350	small nuclear ribonuclear polypeptide C	17.3	MPKFYCDYCDTYLTHDSPSVRK THCSGRKHKE.N	Glu-C
CAI22233	60S acidic ribosomal protein PO	27.5	MPKSKRDKKVSLTKTAKKGLE.L	Glu-C
CAA43925	ribosomal protein L7a	30.0	MPKGGKAKGKKVAPAPAVVKK QE.A	Glu-C
AAH01282	High mobility group nucleosomal binding domain 4	9.5	MPKRKAKGDAKGDKAKVK.D	Asp-N

Table 4.2 N-terminal peptides generated from proteolysis of probable NRMT substrates. Accession numbers, molecular weight and amino acid sequence information derived from GenBank and PubMed entries (NCBI). Optimal human peptide sequences resulting from digestion using endoproteinases Glu-C and Asp-N are listed. Residues C-terminal to expected cleavage sites are listed in X.X format.

Once endoproteinases Glu-C and Asp-N were chosen for proteolysis, another issue became evident. Because modification by NRMT only affects one peptide per protein, detection of this PTM is made more difficult due to high sample complexity following proteolysis of all proteins present. Enrichment of N-terminal peptides therefore provides the optimal conditions for detection. By exploiting the interaction between the anti-me³-SPK antibody and N-terminally modified peptides, we were able to isolate N-terminal peptides only. This afforded the ability to perform two levels of protein characterization. First, the identification of immunoprecipitated proteins was made possible by analyzing non-N-terminal

peptides. Second, analysis of N-terminally modified peptides was used to prove N-terminal methylation of proteins previously identified.

MWCO filters were integral to the separation of both peptide groups. These filters allow only solvated molecular species smaller than the cross-sectional diameter of filter pores to pass through. In this way, it is possible to retain large peptides or proteins on the filter while smaller peptides are collected below. The following methods are illustrated in **Figure 4.6**. Reduction, carbamidomethylation and proteolytic digestion of all immunoprecipitated proteins occurs on top of the MWCO filter. Proteolysis results in Asp-N and Glu-C-generated peptides in solution. The application of centrifugal force to these peptides draws them through the filter while antibodies, endoproteases, extremely large peptides and N-terminally modified peptides remain. Substrates that exhibit weak interactions with antibodies are likely to follow the unmodified peptides. Once unmodified peptides are forced through the filter, a data dependent HPLC-MS/MS analysis that implements CAD and/or ETD MS provides protein identification. Although this analysis ideally does not detect any N-terminal modifications, the identified source protein sequences can be interrogated to decide if N-terminal sequences might promote methylation by NRMT. Using this information, targeted MS/MS scans can be used to ensure that peptides of interest are analyzed.

Disruption of electrostatic antibody:peptide interactions that are used to isolate modified peptides is achieved using a low pH Pierce elution Buffer (pH 2.8).

Competition-based peptide elution is disadvantageous due to the likelihood that

these substrates contribute higher sample complexity in the N-terminal enrichment.

Once N-terminal peptides are eluted, centrifugation draws them into a separate tube

while antibodies, endoproteinases and undigested materials are retained. CAD and

ETD mass analyses were used to analyze the elutions due to variability in length and

charge.

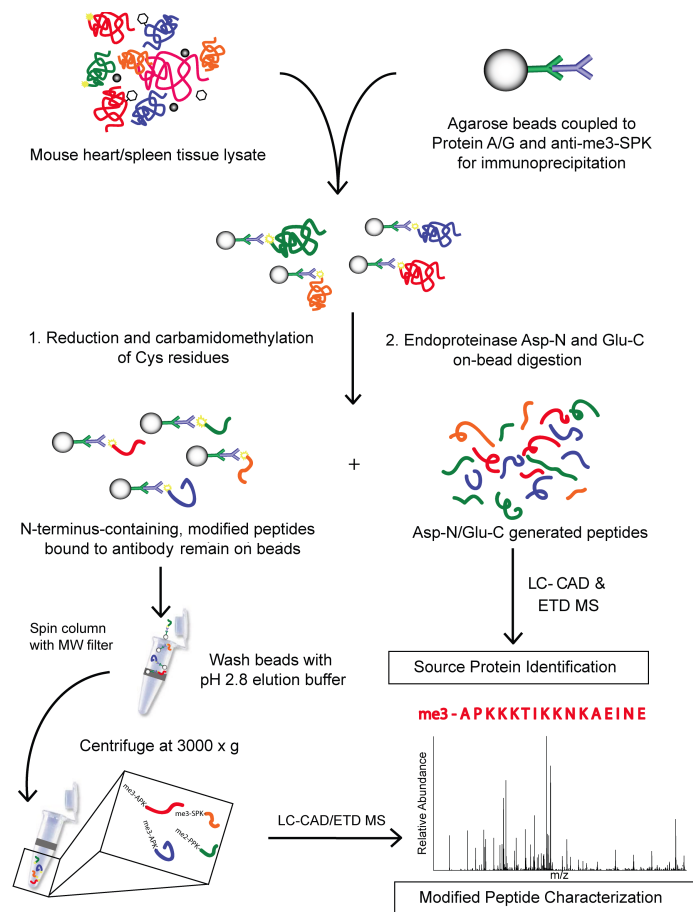


Figure 4.6 Novel strategy for enrichment of N-terminal peptides using methylation-specific antibodies and MWCO filters. Isolation of non-N-terminal peptides first facilitates source protein identification. Selective enrichment of N-terminally methylated peptides provides low sample complexity for efficient HPLC-MS/MS analysis of post-translationally modified peptides. See text for detailed descriptions.

4.5.3.1 Protein Identification by Analysis of Non-N-terminal Peptides

ETD analysis of unmodified peptides isolated by the first step of enrichment resulted in a preliminary identification of hundreds of proteins. The vast majority of identified proteins did not contain the PK, APK, PPK or SPK N-terminal sequences expected to be methylated. We suspect that these non-specifically interact with antibody and Protein A/G during purification. However, several identified peptides did belong to mouse proteins that contain NRMT consensus sequences. These are discussed later as novel NRMT substrates.

We initially expected all methylated N-terminal sequences to contain APK, PPK, or SPK, but nevertheless chose OMSSA search parameters that provided the possibility for any N-terminus to be methylated. The presence of various cytoskeletal proteins in the heart and spleen tissue suggested many non-specific interactions. As expected, we did not detect any N-terminal methylated peptides in the first heart analysis. However, we did identify several N-terminally methylated peptides in the corresponding spleen analysis. Unexpectedly, these peptides have sequences outside of the expected consensus. The first N-terminally methylated peptide identified corresponds to the N-terminus of MLC, regulatory B-like (Accession NP_080340, NCBI). The sequence me3-SSKRAKTKTTKKRPQRATSNVFMFDQSQIQEFKE was confirmed using ETD (**Figure 4.7**). This peptide contains a missed Glu-C cleavage site. The correctly cleaved form was also identified, but the MS/MS spectrum of the missed cleavage provides more comprehensive z•-ion coverage due to the charge on the C-terminal Lys.

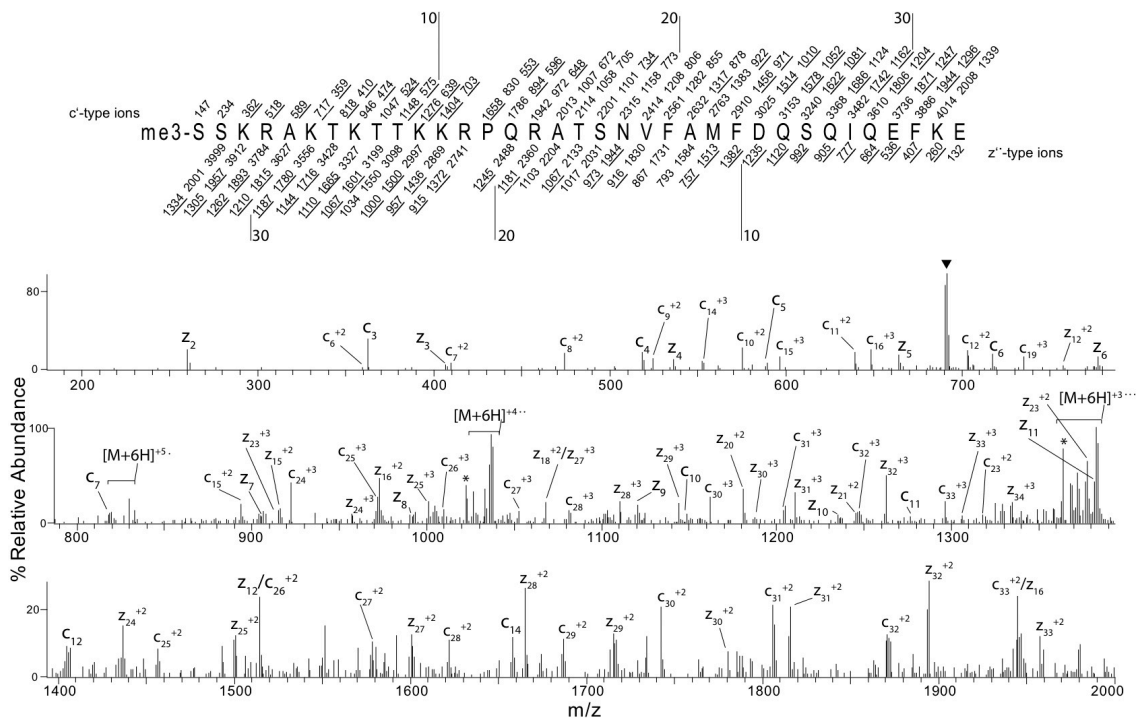


Figure 4.7 ETD MS/MS spectrum of N-terminally trimethylated MLC, regulatory B-like (Accession NP_080340, NCBI) N-terminal peptide SSKRAKTKTKKRPQRATSNVFAMFDQSQIQEFKE. Fragment ions detected are labeled within the spectrum and underlined in the peptide sequence. The inverted triangle corresponds to ions that fall within the prescribed isolation window.

We did not anticipate methylation on this protein because it lacks Pro at residue 2. The discovery of a second me3-SSK peptide was granted by manual inspection of MS¹ data and MS/MS confirmation without a spectrum hit from OMSSA (**Figure 4.8**). The second peptide, which is nearly identical to the first MLC, corresponds to the N-terminal peptide from MLC 12B, regulatory (Accession NP_075891, NCBI). This isoform contains a Lys instead of Arg at position 4 to yield the sequence, me3-SSKAKTKTKKRPQRATSNVFAMFDQSQIQEFKE (MS/MS not shown).

me3-SSKAKTKTTKKRPQRATSNVFAMVDQSIIQEFKE

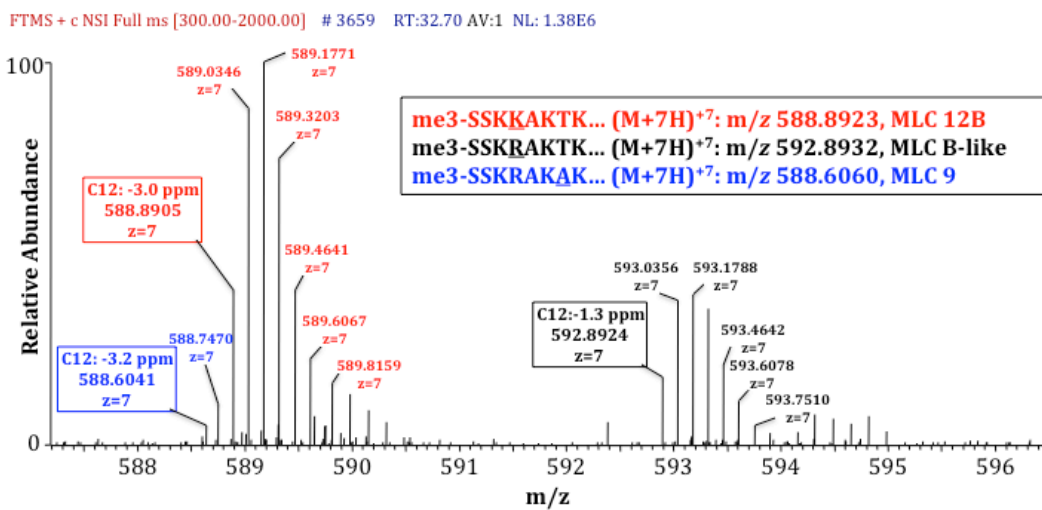


Figure 4.8 Three unique MLCs in varied abundance were observed to contain N-terminal SSK trimethylation. MS/MS analysis of MLC 12B and MLC 9 confirmed these peptide assignments while MLC B-like was only identified with accurate mass measurements.

A third isoform, MLC 9, regulatory (Accession NP_742116) was identified to likely contain N-terminal trimethylation and contained a Thr to Ala substitution at residue 7 (me3-SSKRAKAKTTKKRPQRATSNVFAMFDQSIIQEFKE) (**Figure 4.8**). The two latter isoforms coelute and have nearly identical monoisotopic masses. The co-elution of all MLC forms and low abundance of MLC 9 prevented the acquisition of an unambiguous MS/MS spectrum for this peptide. Identification of MLC9 is only supported by accurate mass measurements from the MS1 scan. While we can't conclude with 100% certainty that this assignment is correct, the presence of the other two isoforms with nearly identical N-terminally modified sequences lends support. These three were the only methylated peptides identified in the protein identification analyses. We believe these were detected in this analysis due to reduced affinity with the anti-me3-SPK antibody. Though it is clear that they are

selectively enriched, we believe that the lack of a Pro2 residue causes reduced affinity. As a result, several washes on the MWCO filter caused elution of these peptides along with the other unmodified peptides. We suspect that because of the low affinity with anti-me3-SPK antibody, inefficient purification of all three methylated N-terminal SSK sequences resulted.

The key observation from these results is that the suspected APK, PPK, SPK sequences are not exclusive substrates for N-terminal methylation. We offer two hypotheses to explain these findings. The simplest is that the NRMT cavity may indeed facilitate interactions with N-terminal sequences without Pro in position 2. This would increase the number of possible NRMT substrates. A second hypothesis is that there may be more than one N-terminal methyltransferase. Preliminary evidence for the discovery of a structurally similar MT, initially named NRMT2, from the Macara lab supports the latter hypothesis (unpublished, data not shown). We suspect that future characterization of NRMT, NRMT2 and any other unidentified N-terminal methyltransferases will better address these findings.

The biological function of MLC methylation is currently unclear. Previous efforts that identified MLC methylation also stated that function of this PTM was unknown (12, 21). These also identified MLC N-terminal trimethylation on alanine only. Our results demonstrate the first discovery of trimethylserine on MLCs. Site directed mutagenesis is best suited to begin to probe functionality of this PTM. Without these studies, we postulate that the modification may be necessary for

interactions with myosin heavy chains (MHCs), motor function, or to block

alternative N-terminal modifications.

4.5.3.2 CAD & ETD Analysis of N-terminal Peptides

The addition of endoproteinase Asp-N and Glu-C digests provided a single HPLC-MS/MS analysis for each tissue. A CAD/ETD toggle method was implemented to interrogate each peptide chosen for MS/MS analysis. Using these methods to analyze the enriched samples, we identified several N-terminally methylated peptides (**Table 4.3**). All peptides were relatively basic ($z > +2$) which required ETD fragmentation for unambiguous identification. While we originally estimated that proteins were originally present at ~ 1 pmol based on the Coomassie stained gel, the estimated abundances of each protein following MS analysis falls within the fmol – pmol range. These results demonstrate the success of our technique and also that we do not lose large amounts of material during enrichment.

Peptide Sequence	Accession	Protein Name	Heart	Spleen
me3-APKKKTINKKNKAE INE.M	NP_766513	kelch-like protein 31	210 fmol	ND
me3-APKKPEPKDDAKA AAPK.A	NP_034989	MLC 3	14 fmol	ND
me2-PPKRIAKRRSPPE.D	NP_598639	regulator of chromosome condensation 1	2 fmol	10 fmol
me3-APKAKKEAPAPPKA E.A	NP_997406	ribosomal protein L23a	0.5 fmol	12 fmol
me3-APKRSAILPQPKKPR PAAAPKLE.D	NP_076360	SET translocation	8 fmol	2.86 pmol/ 38 fmol*
me3-APKKAKKRIEGGSSN VFSMFE.Q	NP_034991	MLC 2, regulatory, cardiac, slow	1.17 pmol	ND
me3- SSKRAKTKTKKRPQ RATSNVFAMFDQSQIQEF KE.A	NP_080340	MLC, regulatory B- like	ND	22 fmol*

me3- SSKKAKTKTTKKRPQ RATSNVFAMFDQSQIQEF KE.A	NP_075891	MLC 12B, regulatory	ND	52 fmol*
me3- SSKRAKAKTKTTKKRPQ RATSNVFAMFDQSQIQEF KE.A	NP_742116	MLC 9, regulatory	ND	5 fmol*

Table 4.3. Summary of peptides and source proteins identified with N-terminal di- or trimethylation. Approximate molar amounts are provided for each protein (ND: Not detected). Molar amounts denoted with * correspond to those from the initial spleen protein identification MS analysis, not the enriched acid elution.

Regulator of Chromatin Condensation (RCC1), NP_598639

RCC1, the first protein shown to require α -methylation for function, was identified in this study as well (**Figure 4.9**). Identification of this peptide suggests that the anti-me3-SPK antibody recognizes the mouse RCC1 me2-PPK N-terminus. The low concentration of RCC1 detected suggests one of several situations is occurring: 1) RCC1 expression levels are naturally low in lysed mouse tissue, 2) RCC1 is not dimethylated at 100% *in vivo*, or 3) lower affinity interactions by me2-PPK with anti-me3-SPK results in low purification yields. We assume that the third effect was a significant contributor due to the structural differences between me2-PPK and me3-SPK. Regardless, N-terminal methylation of human RCC1 has been shown to be essential for chromatin binding, proper spindle pole orientation and normal mitosis (14). Mouse RCC1 likely serves a similar role.

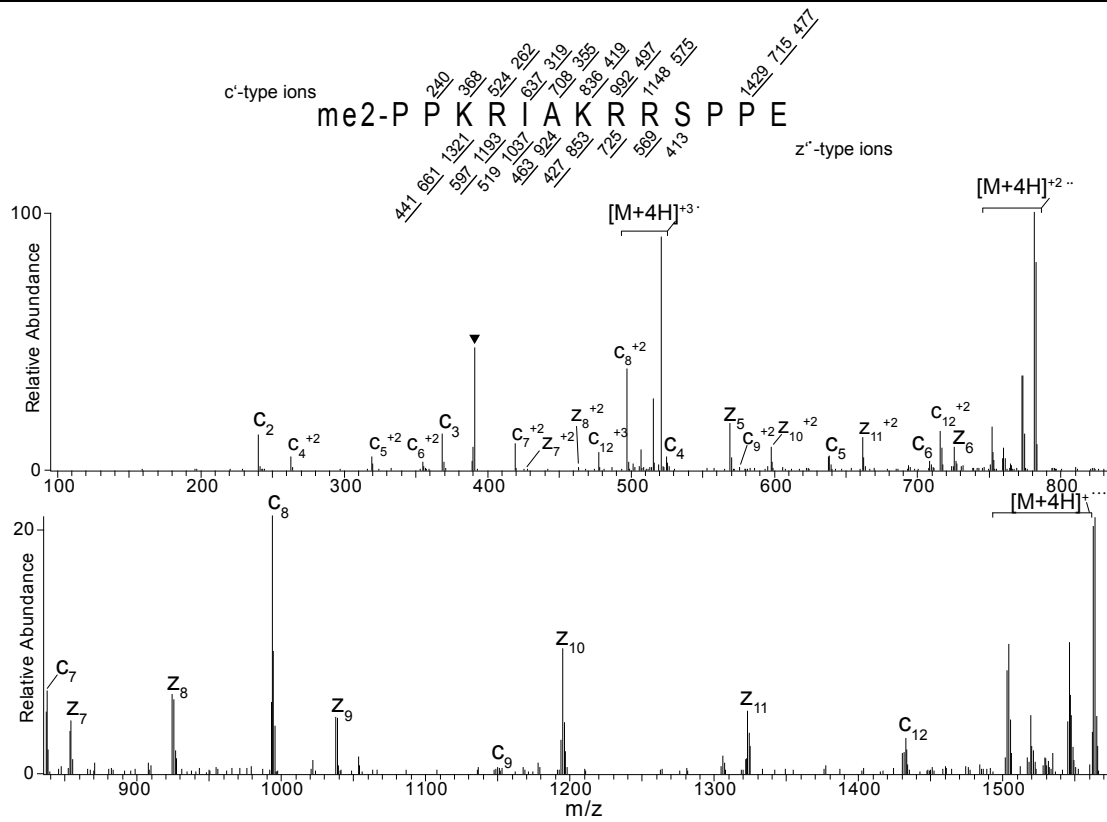


Figure 4.9 ETD MS/MS spectrum of the N-terminal peptide of mouse RCC1 modified with N-terminal dimethylation. Fragment ions detected within the spectrum are labeled and underlined within the peptide sequence. The inverted triangle corresponds to ions that fall within the prescribed isolation window. Neutral losses from the labeled reduced charged species are bracketed.

SET/TAF-1/I2PP2A/INHAT, Accession NP_076360

We identified the trimethylated N-terminus of mouse SET, also referred to as template-activating factor 1 (TAF-1), phosphatase 2A inhibitor I2PP2A (I2PP2A) and inhibitor of histone acetyltransferases (INHAT) (**Figure 4.10**) (22). Human SET was previously identified as an NRMT substrate by Tara Muratore-Schroeder using an earlier anti-me3-SPK antibody IP (18). Therefore, it is unsurprising that we also identified this protein. Detection of SET and RCC1 demonstrates the reproducibility

of this enrichment approach since both proteins were identified in previous anti-me3-SPK IPs.

SET/I2PPA/TAF-1/INHAT has surprisingly diverse functions that includes a role as a histone chaperone protein (22). SET also has known roles in transcription, DNA replication, nucleosome assembly, regulation of acetylation on histone tail side chains and exhibits kinase and phosphatase activity (22-24). The many roles of SET make it difficult to hypothesize a single function for N-terminal methylation, however, it is clear that SET exhibits nuclear localization and likely interacts with DNA. We suspect that the N-terminal tail electrostatically favors interactions with DNA to regulate replication, transcription, nucleosome assembly and other chromatin-associated functions.

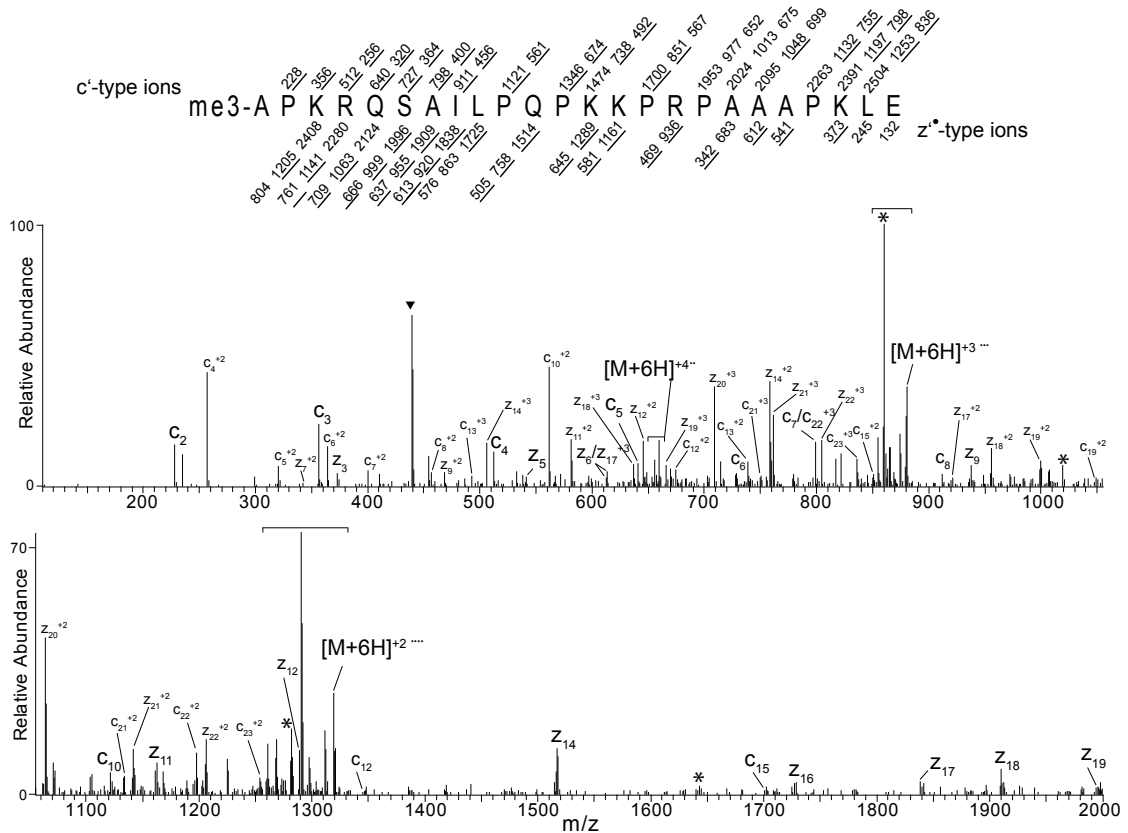


Figure 4.10 ETD MS/MS spectrum of the N-terminal peptide of SET modified with N-terminal trimethylation. Fragment ions detected within the spectrum are labeled and underlined within the peptide sequence. The inverted triangle corresponds to ions that fall within the prescribed isolation window. Ions labeled with an asterisk correspond to losses of trimethylamine. Neutral losses from the labeled reduced charged species are bracketed.

Ribosomal protein L23a, Accession NP_997406

The identification of N-terminally trimethylated ribosomal protein L23a (RL23) (**Figure 4.11**) demonstrates NRMT activity on RNA-associated proteins and suggests evolutionary conservation of the modification since trimethylalanine exists in some *E. coli* ribosomal proteins (25). RL23 is highly conserved in nearly all kingdoms and is located at the ribosome exit tunnel where newly synthesized depart (26). Directly following protrusion from the ribosomal exit tunnel, several important events take place including excision of the initiating Met, chaperone-

facilitated protein folding and N-terminal acetylation. RL23 is known to facilitate ribosomal interactions through binding the chaperone protein Trigger Factor and RNA (27). RL23 is involved in specific interactions with ribosomes and protein processing factors (e.g. signal recognition particle and translocon) required for targeting of newly synthesized proteins to the endoplasmic reticulum for processing (26). Based on the reports that demonstrate RL23 interactions with ribosomal RNA, we suspect that N-terminal methylated RL23a enhances binding with negatively charged ribosomal RNA by placing a fixed positive charge at the N-terminus.

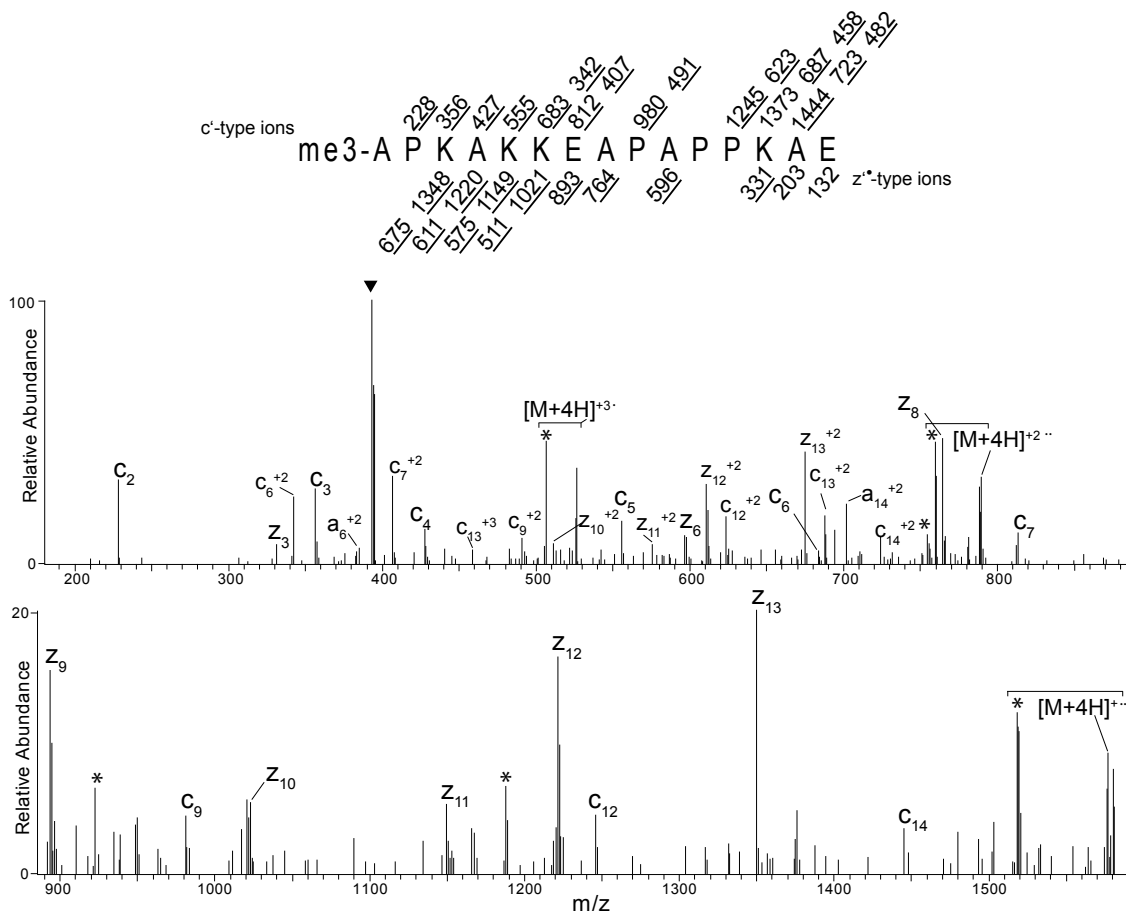


Figure 4.11 ETD MS/MS spectrum of the N-terminally trimethylated peptide from mouse ribosomal protein L23a. Fragment ions detected within the spectrum are labeled and underlined within the peptide

sequence. The inverted triangle corresponds to ions that fall within the prescribed isolation window. Ions labeled with an asterisk correspond to losses of trimethylamine. Neutral losses from the labeled reduced charged species are bracketed.

Kelch-like protein 31, Accession NP_766513

Kelch-like proteins contain multiple kelch motifs which are 44-56 amino acid stretches that form beta sheets (28, 29). Repetition of 4 to 7 kelch motifs can form a β -propeller through which protein interactions are mediated. The identification of N-terminal trimethylation on kelch-like protein 31 (KLHL31) (**Figure 4.12**) contributes a novel PTM to a relatively uncharacterized protein that stems from a family with incredibly diverse functions (28, 29). Studies of mouse KLHL31 are sparse. Human KLHL31 was found to be highly expressed in the cytoplasm and nucleus of embryonic skeletal and heart muscle; this agrees with the detection of KLHL31 in heart tissue in our study (29). Human KLHL31 was implicated as an inhibitor to transcriptional activities of serum and TPA-response elements and affected MAPK signaling by reducing JNK-mediated phosphorylation of c-Jun (29). Yu et al. hypothesize that regulation of transcription by MAPK signalling is in part afforded by KLHL31 interaction with other proteins to control heart development. We speculate that mouse KLHL31 interacts similarly. We also suspect that the fixed charge resulting from N-terminal methylation regulates the electrostatic interactions between KLHL31 and other proteins or DNA during transcription. This is supported by data that shows kelch motif-containing proteins have the ability to recognize specific DNA sequences (28).

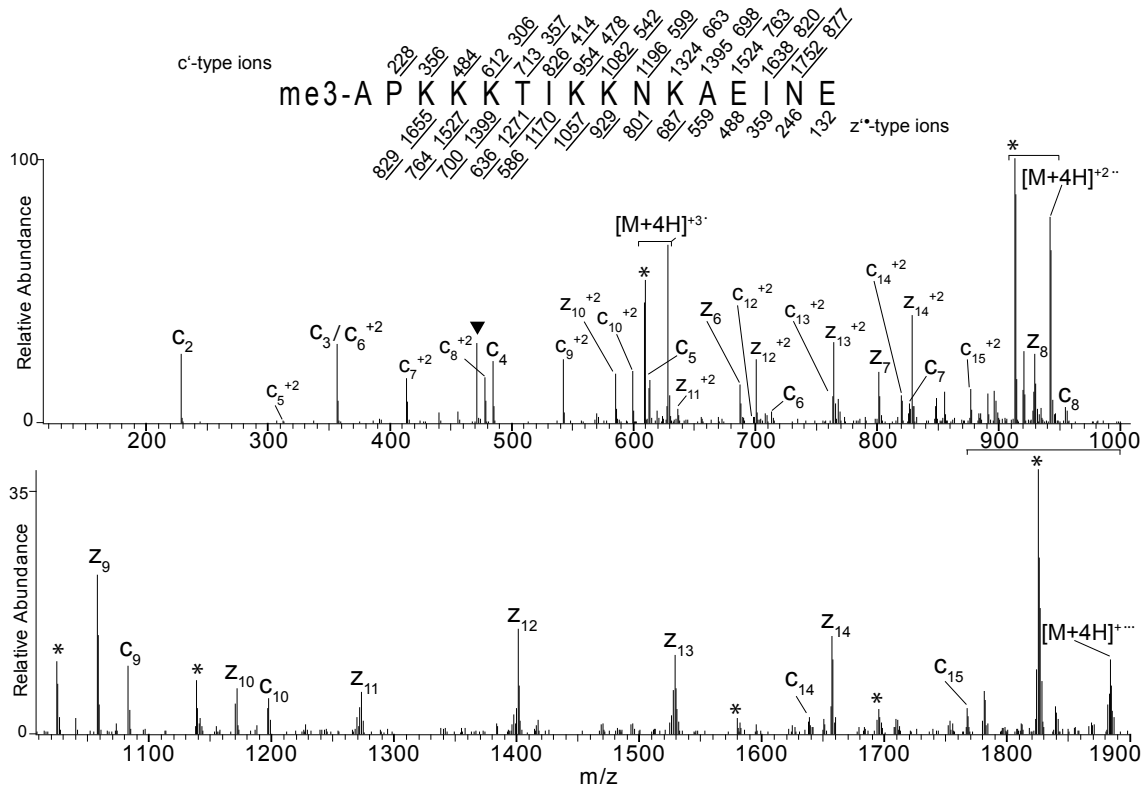


Figure 4.12 ETD MS/MS spectrum of the N-terminal peptide of kelch-like protein 31 modified with N-terminal trimethylation. Fragment ions detected within the spectrum are labeled and underlined within the peptide sequence. The inverted triangle corresponds to ions that fall within the prescribed isolation window. Ions labeled with an asterisk correspond to losses of trimethylamine. Neutral losses from the labeled reduced charged species are bracketed.

MLC 3, Accession NP_034989 & MLC 2, Accession NP_034991

Two additional MLCs were identified as NRMT substrates that exhibit trimethylated N-terminal APK sequences (**Figures 4.13, 4.14**). Because we identified three me3-SSK MLCs and rabbit MLCs were previously known to be methylated, it seems that that N-terminal methylation is conserved across MLC types but also across species (12, 21). N-terminal methylation of MLC residues following Met cleavage appears to be evolutionary conserved so we suspect that this

PTM is essential for MLC binding to MHCs and/or proper motor function. Site

directed mutagenic studies should be able to address this further.

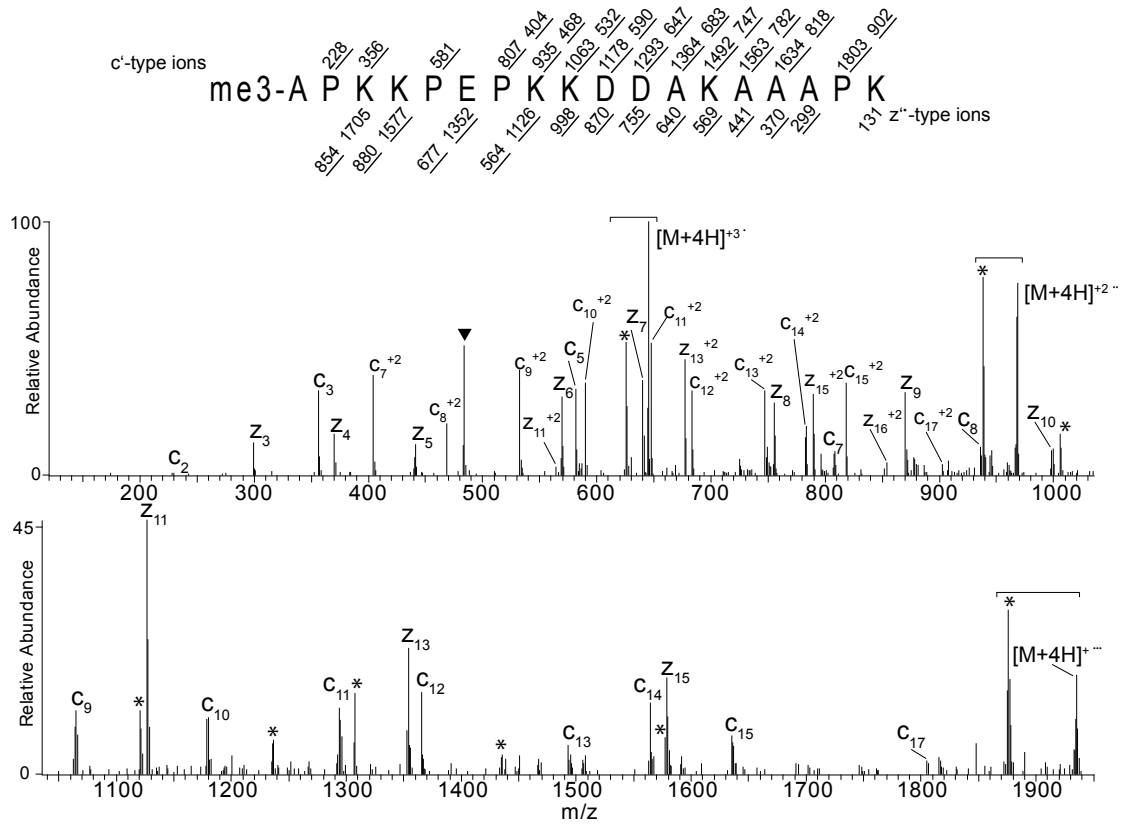


Figure 4.13 ETD MS/MS spectrum of the N-terminal peptide of myosin light polypeptide 3 modified with N-terminal trimethylation. Fragment ions detected within the spectrum are labeled and underlined within the peptide sequence. The inverted triangle corresponds to ions that fall within the prescribed isolation window. Ions labeled with an asterisk correspond to losses of trimethylamine. Neutral losses from the labeled reduced charged species are bracketed.

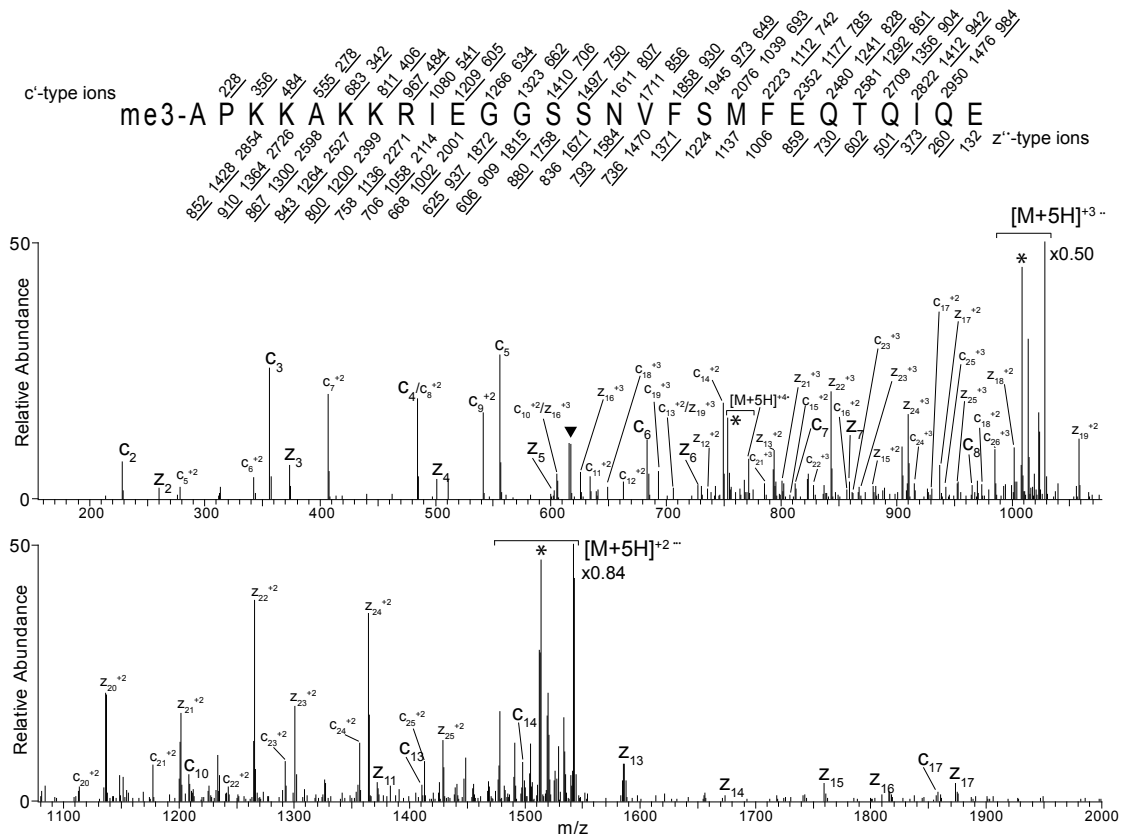


Figure 4.14 ETD MS/MS spectrum of the N-terminal peptide of myosin light polypeptide 2 modified with N-terminal trimethylation. Fragment ions detected within the spectrum are labeled and underlined within the peptide sequence. The inverted triangle corresponds to ions that fall within the prescribed isolation window. Ions labeled with an asterisk correspond to losses of trimethylamine. Neutral losses from the labeled reduced charged species are bracketed.

Taken together, these data provide the identification of several novel α -N-methylation substrates, which we suspect result from the enzymatic action of NRMT. SDS-PAGE analysis of proteins purified using the anti-me3-SPK antibody shows three prominent bands (**Figure 4.15A**). Immunoblotting against anti-me3-SPK reveals the same three bands and an additional one near 20 kDa (**Figure 4.15B**). Based on the molecular weights of the proteins we identified, we attribute these bands to the following proteins: kelch-like protein 31, RCC1, SET, and the isobaric MLCs and RL23.

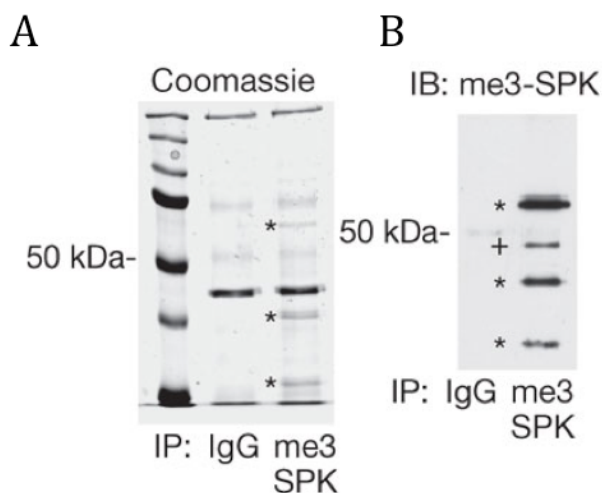


Figure 4.15 SDS-PAGE and Western blot analysis of anti-me3-SPK immunopurified proteins. (A) Coomassie stained, SDS-PAGE separated proteins from an anti-me3-SPK IP results in several protein of varied molecular weight (*). (B) Immunoblotting with anti-me3-SPK results in identical bands with an additional band near 20 kDa (+). The proteins identified as NRMT substrates in this study correlate to these bands: kelch-like 31 (70 kDa), RCC1 (45 kDa), SET (33 kDa), and MLCs or RL23 (20 kDa).

4.5.4 Signature Loss of Trimethylamine Under ETD Fragmentation

Upon inspection of ETD spectra from several peptides, ions were identified that were ~ 60 Da less than singly charged c-ions and reduced charge species (denoted by * in above spectra). Further examination of the data showed that these ions were only present under ETD fragmentation of trimethylated N-termini. In addition, losses were more prevalent from heavier c-ions and reduced charge species; neutral losses from these ions corresponded to a difference of 58 Da and 59 Da, respectively. Similar dissociation has been reported to occur from precursor ions under CAD (30). We propose that these ions result following capture of a second electron during ETD. Radical-directed fragmentation by a second electron transfer would cause cleavage at the N-terminal fixed charge to result in loss of trimethylamine (59 Da). Ions that correspond to the reduced charge species minus

59 Da are often more abundant than the reduced charge species itself and support this type of fragmentation (**Figures 4.10-4.14**). Ions that correspond to one Da less than the loss of trimethylamine (-58 Da) suggest dissociation involving c-ions of a higher charge state that carry multiple H^+ (+1 Da each). This is supported by the observation that these ions are only evident near higher mass c-ions, which have the propensity to carry more charge. An example of the proposed fragmentation that generates these ions is illustrated in **Figure 4.16**. We assume this multiple electron transfer process occurs because the N-terminal peptides are highly basic and all include an N-terminal fixed charge, which increases the overall electron affinity for each ion.

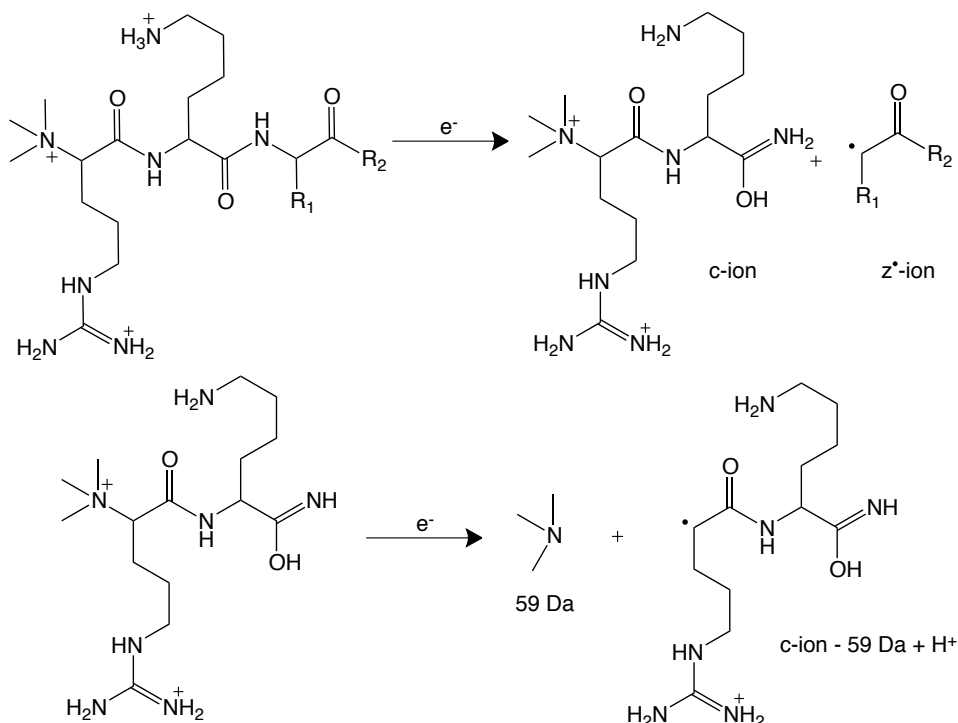


Figure 4.16 Secondary electron transfer to multiply charged c-ions directs fragmentation at the N-terminal trimethylamine. The resulting “c-like ion” has a mass equal to the intact c-ion -58 Da. Identical ions are generated from additional electron capture from reduced charge species that lose trimethylamine first and

then fragment to form c- and z \bullet -type ions. Similar chemistry causes loss of trimethylamine (- 59 Da) from reduced charge species.

The presence of this characteristic loss is helpful for the identification of N-terminally methylated substrates. Since Pro residues exist at position 2 in the majority of identified NRMT substrates, we observe a lack of c- and z \bullet -ions at the extreme N-terminus due to the inability of a single electron transfer to cause cleavage of both N-C bonds at the tertiary nitrogen of Pro (31). These signature ion losses can be used to add confidence to the presence of α -N-methylation when c₁ and the largest z-ion are not present, because trimethylamine losses are uniquely abundant with N-terminal methylation.

4.6 Conclusions

The data presented in this chapter support new methods for the identification of N-terminally methylated proteins. These are built on the ability to IP using a me₃-SPK specific antibody and exploit MWCO filters in order to selectively isolate two groups of peptides for analysis by MS. These methods reduce the amount of time spent on individual preparation steps including Cys reduction, carbamidomethylation and proteolytic digestion by simultaneously performing all steps in batch prior to separation of both peptide groups. The first group is composed of proteolyzed proteins without extreme N-terminal peptides. Analysis of these peptides using conventional HPLC-MS/MS techniques provides identification of abundant proteins isolated using IP. By researching the N-terminal sequences that belong to these proteins, we gain insight into those that are

potentially non-specifically purified. The knowledge of proteins that are present and exhibit NRMT consensus sequences can be used to target potentially modified peptides for MS/MS. The second peptide group is comprised of N-terminal peptides following antibody elution and centrifugation. These peptides are free of contaminating antibodies, endoproteinases and unmodified peptides and provide optimal conditions for identifying α -N-methylation present at low abundance.

Our results provide the first list of enriched mammalian N-terminally modified proteins identified using mass spectrometry. The list represents the following N-terminally methylated proteins in either spleen and/or heart mouse tissue: RCC1, SET, MLC regulatory B-like, MLC 12B, MLC 9 regulatory, ribosomal protein L23a, kelch-like protein 31, MLC 3 and MLC 2 regulatory, cardiac, slow. Initially, sequences known to accept N-terminal methylation in higher eukaryotes was limited to mainly APK, PPK and SPK. However, identification of three unique myosin proteins (MLC regulatory B-like, MLC 12B and MLC 9) challenges this thought by demonstrating that N-terminal SSK sequences are also substrates.

NRMT, the methyltransferase discovered to modify RCC1, SET and retinoblastoma protein, is suspected to be the enzyme that catalyzes methylation of these proteins. It remains possible, however, that other unidentified mammalian N-terminal methyltransferases exist to broaden the availability of modifiable proteins within the proteome. Discovery of another possible NRMT-like enzyme with

structural homology tentatively named NRMT2 in Dr. Macara's laboratory supports this hypothesis.

Perhaps the most curious part of α -N-methylation is its natural abundance. Since no previous methods have characterized this PTM, it is difficult to estimate the prevalence of N-terminal methylation throughout the proteome. Estimates that state over 80% of all proteins are N-acetylated are likely inaccurate because many N-acetylation studies were performed using low resolution MS (7). Since N-terminal acetylation and trimethylation modifications are nearly isobaric, it would be easy to subscribe any addition of 42 Da at the N-terminus to acetylation. We suspect that many N-acetylated proteins may actually be α -N-methylated.

The N-terminally methylated proteins we identified vary in function and include ribosomal proteins, motor proteins, and those associated with mitosis. Evidently, N-terminal methylation is not limited to certain protein classes, but is instead ubiquitous. Because several identified proteins have uncharacterized function, it is difficult to postulate the role of N-terminal methylation for each. Because Chen et al. provide the only known function of α -N-methylation on RCC1 (14), we need to use their results as the basis for our speculation of PTM function for these newly identified proteins. Our study opens up a new field of PTM characterization using mass spectrometric detection for novel NRMT substrates. In the future, we expect to see more studies on N-terminal methylation due to the critical function of N-terminally methylated RCC1 and the recent identification of

NRMT. We believe that the HPLC-MS/MS methods described within this chapter

will be highly beneficial for these studies.

4.7 References

1. Farley, A. R. & Link, A. J. (2009) in *Methods in Enzymology*, ed. Deutscher, R. R. B. a. M. P. Academic Press: 725-763.
 2. Arginine Methylation: An Emerging Regulator of Protein Function, Bedford, M. T. & Richard, S., *Molecular Cell* **2005** (18): 263-272.
 3. Post-translational modifications in signal integration, Deribe, Y. L., Pawson, T. & Dikic, I., *Nature Structural & Molecular Biology* **2010** (17): 666-672.
 4. Posttranslational Modification of Proteins: Expanding Nature's Inventory, Walsh, C. T. **2006**, Roberts and Company Publishers, Englewood, CO: 490.
 5. Extent of N-terminal modifications in cytosolic proteins from eukaryotes, Martinez, A., Traverso, J. A., Valot, B., Ferro, M., Espagne, C., Ephritikhine, G., Zivy, M., Giglione, C. & Meinel, T., *Proteomics* **2008** (8): 2809-2831.
 6. Methionine or not methionine at the beginning of a protein, Sherman, F., Stewart, J. W. & Tsunasawa, S., *BioEssays* **1985** (3): 27-31.
 7. N-terminal Acetyltransferases and Sequence Requirements for N-terminal Acetylation of Eukaryotic Proteins, Polevoda, B. & Sherman, F., *Journal of Molecular Biology* **2003** (325): 595-622.
 8. The primary structure of protein L16 located at the peptidyltransferase center of Escherichia coli ribosomes, Brosius, J. & Chen, R., *FEBS Letters* **1976** (68): 105-109.
 9. Primary structure of protein L33 from the large subunit of the Escherichia coli ribosome, Wittmann-Liebold, B. & Pannenbecker, R., *FEBS Letters* **1976** (68): 115-118.
 10. Novel N-terminal protein blocking group identified as dimethylproline, Pettigrew, G. W. & Smith, G. M., **1977** (265): 661-662.
 11. N-terminal methylation of proteins: Structure, function and specificity, Stock, A., Clarke, S., Clarke, C. & Stock, J., *FEBS Letters* **1987** (220): 8-14.
 12. The occurrence of [alpha]-N-trimethylalanine as the N-terminal amino acid of some myosin light chains, Henry, G. D., Dalgarno, D. C., Marcus, G., Scott, M., Levine, B. A. & Trayer, I. P., *FEBS Letters* **1982** (144): 11-15.
 13. Regulation of chromatin binding by a conformational switch in the tail of the Ran exchange factor RCC1, Hao, Y. & Macara, I. G., *The Journal of Cell Biology* **2008** (182): 827-836.
 14. N-terminal [alpha]-methylation of RCC1 is necessary for stable chromatin association and normal mitosis, Chen, T., Muratore, T. L., Schaner-Tooley, C. E., Shabanowitz, J., Hunt, D. F. & Macara, I. G., *Nature Cell Biology* **2007** (9): 596-603.
 15. Structure of RCC1 chromatin factor bound to the nucleosome core particle, Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S., **2010** (467): 562-566.
 16. The Ran GTPase as a marker of chromosome position in spindle formation and nuclear envelope assembly, Hetzer, M., Gruss, O. J. & Mattaj, I. W., **2002** (4): E177-E184.
-

17. Chromatin Docking and Exchange Activity Enhancement of RCC1 by Histones H2A and H2B, Nemergut, M. E., Mizzen, C. A., Stukenberg, T., Allis, C. D. & Macara, I. G., Science **2001** (292): 1540-1543.
 18. NRMT is an N-methyltransferase that methylates RCC1 and retinoblastoma protein, Schaner Tooley, C. E., Petkowski, J. J., Muratore-Schroeder, T. L., Balsbaugh, J. L., Shabanowitz, J., Sabat, M., Minor, W., Hunt, D. F. & Macara, I. G., Nature **2010** (466): 1125-1128.
 19. A Neutral Loss Activation Method for Improved Phosphopeptide Sequence Analysis by Quadrupole Ion Trap Mass Spectrometry, Schroeder, M. J., Shabanowitz, J., Schwartz, J. C., Hunt, D. F. & Coon, J. J., Analytical Chemistry **2004** (76): 3590-3598.
 20. Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry, Udeshi, N. D., Compton, P. D., Shabanowitz, J., Hunt, D. F. & Rose, K. L., Nature Protocols **2008** (3): 1709-1717.
 21. The widespread distribution of [alpha]-N-trimethylalanine as the N-terminal amino acid of light chains from vertebrate striated muscle myosins, Henry, G. D., Trayer, I. P., Brewer, S. & Levine, B. A., European Journal of Biochemistry **1985** (148): 75-82.
 22. Identification of distinct SET/TAF-Ibeta domains required for core histone binding and quantitative characterisation of the interaction, Karetsou, Z., Emmanouilidou, A., Sanidas, I., Liokatis, S., Nikolakaki, E., Politou, A. & Papamarcaki, T., BMC Biochemistry **2009** (10): 1-12.
 23. The nuclear localization of SET mediated by impa3/impb attenuates its cytosolic toxicity in neurons, Qu, D., Zhang, Y., Ma, J., Guo, K., Li, R., Yin, Y., Cao, X. & Park, D. S., Journal of Neurochemistry **2007** (103): 408-422.
 24. SET and PARP1 remove DEK from chromatin to permit access by the transcription machinery, Gamble, M. J. & Fisher, R. P., **2007** (14): 548-555.
 25. N-trimethylalanine, a novel blocking group, found in E. coli ribosomal protein L11, Lederer, F., Alix, J.-H. & Hayes, D., Biochemical and Biophysical Research Communications **1977** (77): 470-480.
 26. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins, Kramer, G., Boehringer, D., Ban, N. & Bukau, B., **2009** (16): 589-597.
 27. L23 protein functions as a chaperone docking site on the ribosome, Kramer, G., Rauch, T., Rist, W., Vorderwulbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E. & Bukau, B., **2002** (419): 171-174.
 28. The kelch repeat superfamily of proteins: propellers of cell function, Adams, J., Kelso, R. & Cooley, L., Trends in Cell Biology **2000** (10): 17-24.
 29. A novel human BTB-kelch protein KLHL31, strongly expressed in muscle and heart, inhibits transcriptional activities of TRE and SRE, Yu, W., Li, Y., Zhou, Z., Deng, Y., Wang, Z., Yuan, W., Li, D., Zhu, C., Zhao, X., Mo, X., Huang, W., Luo, N.,
-

-
- Yan, Y., Ocorr, K., Bodmer, R., Wang, Y. & Wu, X., Molecules and Cells **2008** 443-453.
30. Differentiation between peptides containing acetylated or tri-methylated lysines by mass spectrometry: An application for determining lysine 9 acetylation and methylation of histone H3, Zhang, K., Yau, P. M., Chandrasekhar, B., New, R., Kondrat, R., Imai, B. S. & Bradbury, M. E., Proteomics **2004** (4): 1-10.
31. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F., Proceedings of the National Academy of Sciences of the United States of America **2004** (101): 9528-9533.
-